Operations Research Symposium

U. S. ARMY ELECTRONICS COMMAND
FORT MONMOUTH, N. J.

29-30 March 1966

Technical Papers - Part 1

OFFICE, CHIEF OF RESEARCH AND DEVELOPMENT
DEPARTMENT OF THE ARMY

DEFENSE SUPPLY AGENCY
HEADQUARTERS, DEFENSE DOCUMENTATION CENTER
FOR SCIENTIFIC AND TECHNICAL INFORMATION
CAMERON STATION
ALEXANDRIA, VIRGINIA 22314

TELEPHONE
AREA CODE 202
OXFORD ~~XXXXX~~
81984

IN REPLY REFER TO DDC-~~XXXX~~-TCA 052467-3378

24 May 1967

SUBJECT: Request for Scientific and Technical Reports

TO: Commanding Officer
Army Research Office (Durham)
Box CM, Duke Station
Durham, North Carolina 27706

1. The report referenced below is believed to be a DoD-funded document. It cannot be located in the Defense Documentation Center (DDC) collection.

2. DoD Instruction 5100.38 of 29 March 1965 requires twenty (20) copies of the report be forwarded to DDC. At least one copy should be black printing on white background (or if typed, the ribbon copy) suitable for reproduction by photographic techniques.

3. Documents must be stamped with proper distribution statements according to DoD Directive 5200.20 of 29 March 1965. This includes information on the releasability of unclassified documents for sale by the Clearinghouse for Scientific and Technical Information.

4. If the document is not releasable to DDC, indicate the category of exception by checking the applicable statement under A on the reverse side of this letter. When reports are forwarded to DDC, or have already been sent to DDC, fill in statements under B on the reverse.

5. Return the copy of this form with the reports or as a reply to this request. Enclosed is a DDC Form 50 for your use, if you wish to be notified of our accession (AD) number. A franked label is also enclosed to defray cost of shipment of documents.

2 Encl.
1. DDC Form 50
2. Franked label

Roy H. Chapman
ROY H. CHAPMAN
Director
User Services

052467-3378

Office, Chief of Rsch & Dev - Dept of Army - Wash, D.C.

Operations Research Symposium - Technical Papers, Pt I,
29-30 March 1966

DDC has part II as AD 370 995.

Note: We are ordering this report direct from you, per information received from Army Electronics Command Fort Monmouth, N.J. stating that you sponsored the Symposium.

GMR 14558 I/S

A. In accordance with the provisions of Department of Defense Instruction 5100.38 of 29 March 1965, the document requested cannot be supplied for the following reasons:

- ☐ It is TOP SECRET.
- ☐ It contains cryptographic and communications security.
- ☐ It is excepted in accordance with the DoD instructions which apply to communications and electronic intelligence.
- ☐ It is a registered document or publication.
- ☐ It is an administrative paper, memorandum or report, a contract or grant proposal, or an order.

B.

- ☒ 20 copies are forwarded herewith.
- ☐ The document requested has already been forwarded to DDC by our letter of ________.
- ☐ The document requested is AD number ________.

JAMES A. SPARKS

Print or Type Name

James A. Sparks

AUTHORIZED SIGNATURE

PREPRINTS

for the

United States Army

OPERATIONS RESEARCH SYMPOSIUM

29-30 March 1966

PART I
(Unclassified Volume)

Sponsored by

Office, Chief of Research and Development
Department of the Army

Hosted and Conducted by
Headquarters, U. S. Army Electronics Command
Fort Monmouth, New Jersey

U. S. ARMY OPERATIONS RESEARCH SYMPOSIUM

29-30 March 1966

FOREWORD

In the last few years the Army has become increasingly more conscious of the employment of modern study techniques in the examination of alternative decisions which may be made concerning current and future operations. With the increasing costs of complex weapons systems and increasing attention to the development of a multipurpose Army, it is vitally necessary that full attention be directed toward the improvement of these study techniques themselves. One method by which this may be accomplished is the conduct of symposia designed to bring the participants abreast of current developments.

The theme of this year's Army Operations Research Symposium is "Life Cycle Management of Materiel." The papers to be presented are compiled in this volume. Being responsive to the theme, they are primarily hardware oriented. As in previous years, papers have been accepted from Army personnel working on operations research projects and from personnel of the several contractors working in this field for the Army.

This compilation of preprints represents a new procedure with respect to this series of symposia. There will be no formal post-symposium proceedings, as there have been in other years. It is believed that the new system will elicit more valuable discussion of the papers at the symposium itself.

The Department of the Army has encouraged this series of symposia as one means of stimulating the in-house operations research capability. The forthcoming symposium evidences progress in this direction since it is observed that of the papers accepted for presentation, approximately three-fourths are by Army operations research personnel.

Robert E. Kimball

ROBERT E. KIMBALL
Colonel, GS
Director of Army Research

PAGES NOT FILMED ARE BLANK

TABLE OF CONTENTS

A STOCHASTIC SURVEILLANCE DECISION MODEL
CYBERNETIC CORRELATORS

by
Erwin Biser
Avionics Laboratory
USAECOM, Ft Monmouth, N. J.

1. INTRODUCTION:

The notion of multisensor surveillance systems has come to the fore in recent years. It is pertinent, therefore, to postulate a surveillance system consisting of a finite number of heterogeneons sensors such as photographic, optical, electromagnetic (types of radar), infrared (IR), lasers, etc. The concepts of organization, in the cybernetic sense, and of information content naturally suggest themselves with regard to problems of system analysis, system synthesis and integration of a multisensor surveillance and reconnaissance complex.

2. SENSOR EVENTS, TARGET EVENTS AND THE SYSTEM CONCEPT:

Associated with each sensor are sensor-events; these are responses (sometimes spurious) of the sensors, say, the blip on the radar scope, to objects "out there" in space which may yield information about the presence or absence of targets. It is by means of the sensor-events that detection identification, and recognition of targets are accomplished. The knowledge of the characteristics of the sensor-events does not determine uniquely and absolutely the presence or absence (let alone the type) of a target. These sensor-data may also take the form of the values of parameter measurements such as the average power of a sampled signal.

By a target event is meant simply the presence or absence of a target "out there" in a spatio-temporal region that corresponds to or <u>is inferred from</u> a set of sensor-events or their parameter values. Note that a sensor is expected to yield a return on the basis of which the presence or absence of a target is estimated. Even in

the case of a radar sensor the range-azimuth-elevation triplet may not indicate the presence or absence of a target with certainty.

3. THE CONCEPT OF INFORMATION CONTENT OF A SYSTEM:

The concept of randomness is the fundamental concept underlying probability, statistics and information theory; it is this concept that forms the inextricably close relationships of the three mathematical disciplines. With entropy is associated uncertainty. Entropy is the quantity that measures the amount of uncertainty (the mean uncertainty) associated with a probability space:

$$1.\quad H(A) = -\sum_{x \in A} p(A) \log p(A)$$

where the probability space is given by:

$$2.\quad A = \begin{pmatrix} A_1, A_2, \ldots, A_n \\ p_1, p_2, \ldots, p_n \end{pmatrix}$$

The logarithms are taken to an arbitrarily selected base(except zero); and $p_i \log p_i = 0$ if $p_i = 0$. $H(p_1, p_2, \ldots, p_n)$ is an increasing function of the p_i's. It is equal to zero if and only if one of the p_i's takes on the value 1 and the others are zero; in all other cases it is always positive. The entropy of a space increases as the probabilities in the space decrease.

H(A), the entropy of A, can also be interpreted as yielding the mean uncertainty about A after an observation; and as providing the mean uncertainty about the probability space A prior to an observation. The terms uncertainty and information will be used interchangeably.

It can be shown that the entropy of the product space (X Y) is given by:

$$3.\quad H(XY) = -\sum_{XY} p(x,y) \log p(x,y)$$

$$4.\quad H(XY) = H(X) + H(Y|X)$$

where $H(Y|X)$ measures the average additional amount of information needed to specify the values of the elements of the probability space Y if the values of the events of the space X are known. $H(Y|X)$ is the conditional entropy of the space Y given that the events of the space X have been observed.

It can also be shown that:

$$4. \quad H(Y|X) \leq H(Y)$$

This is Shannon's Fundamental Inequality.

This equation states that the entropy of a finite probability space Y cannot be increased as a result of observations on space X . It can also be shown that:

$$5. \quad H(XY) \leq H(X) + H(Y)$$

The equality sign holds if and only if the spaces X and Y are statistically independent. This inequality can be generalized to the product of any finite number of spaces:

$$6. \quad H(XYZ) \leq H(X) + H(Y) + H(Z)$$

The last inequality constitutes the condition for the information space (XYZ) to be consistent. It has been shown that *

$$7. \quad H(S) \leq H(S_{i_1}) + H(S_{i_2}) + \ldots + H(S_{i_K})$$

*Erwin Biser, "Partitions of Discrete Information Spaces With Some Systems Applications." Transactions of the Tenth Conference of Army Mathematicians; June, 1965, ARO-D 65-2.

where $\{S_{\mu_1}, S_{\mu_2}, \ldots, S_{\mu_k}\}$ is a k-partition of S (a probability space). A k-partition of S, symbolized by $\pi^{(k)}(S)$ is given by the two conditions:

$$8(a) \quad S = S_{\mu_1} \cup S_{\mu_2} \cup, \ldots \cup S_{\mu_k}$$

$$8(b) \quad S_{\mu_j} \cap S_{\mu_k} = \phi$$

ϕ is (the null set), for $j \neq k$

This means that the information space S is the union of a finite number of subspaces that are pairwise disjoint, where S_{μ_1} (likewise S_{μ_2}) is a subspace of S, that is by itself a product space of μ_j spaces.

It can be shown that if $\{S_{\mu_1}, S_{\mu_2}\}$ is a 2-partition of an information space S, then:

$$9. \quad H(S_{\mu_1} | S_{\mu_2}) = H(S) - H(S_{\mu_2})$$

$$10. \quad H(S_{\mu_2} | S_{\mu_1}) = H(S) - H(S_{\mu_1})$$

$H(S_{\mu_1} | S_{\mu_2})$ is the amount of information provided by subspace (subsystem) S_{μ_2} about subspace (subsystem) S_{μ_1}. Likewise, for $H(S_{\mu_2} | S_{\mu_1})$

EXAMPLE

$S = (X_1 X_2 X_3 X_4)$, the product space of X_1, X_2, X_3, X_4

$$11. \quad S_{\mu_1} = (X_2 X_4) \; ; \; S_{\mu_2} = (X_1 X_3)$$

$$12. \quad S = S_{\mu_1} \cup S_{\mu_2} \; ; \; S_{\mu_1} \cap S_{\mu_2} = \phi$$

$$13. \quad H(S) = H(X_1 X_2 X_3 X_4)$$

14. $H[(X_2 X_4)/(X_1 X_3)] = H(X_1 X_2 X_3 X_4) - H(X_1 X_3) \geq 0$

The foregoing constitutes the information theoretical concept of system as an information space, as the product space of parameters that are effectively system descriptors. The parameters can stand for operations, activities, functions of subsystems, and probabilities, e.g. probability of detection, probability of identification, etc. For our purposes each parameter is restricted to a finite set of values. Thus the subspaces alluded to in the preceding discussion are essentially subsystems of an information system.

4. THE CONDITIONAL ENTROPY CRITERION:

Let $\mathcal{T}$ be a set or class (pattern) of target events. $\mathcal{T} = \{T_1, T_2\}$ is a 2-partition of $\mathcal{T}$. Similarly, let $\mathcal{S} = \{S_1, S_2\}$ be a 2-partition of $\mathcal{S}$, a set or class or pattern of sensor-events, composed of responses to two types or classes of targets, T_1 and T_2.

What we have in effect is a system responding to two classes of target-events; and yielding two classes of sensor-events.

Let $p(T_i/S_j)$ be the probability that $T \in T_i$ if the observable $S \in S_j$, where $i, j = 1, 2$.

$p(T_i, S_j)$ is the probability that $T \in T_i$ and $S \in S_j$ simultaneously.

$H(T/S)$ = the additional information about given the observation of the sensor-event S_j.

15. $$H(T/S) = \sum_{i=1}^{2}\sum_{j=1}^{2} p(T_i|S_j)\,p(S_j) \cdot \ln\left[\frac{p(T_i|S_1)p(S_1) + p(T_i|S_2)p(S_2)}{p(T_i|S_j)\,p(S_j)}\right]$$

Equation 15 is obtained from the following considerations:

16. $$H(T/S) = -\sum_{T\in T_i}\sum_{S\in S_j} p(T_i, S_j)\ln p(T_i|S_j)$$

17. $$= \sum_{T\in T_i}\sum_{S\in S_j} p(T_i, S_j)\ln\left[\frac{1}{p(T_i|S_j)}\right]$$

18. $$p(S_j|T_i)\,p(T_i) = p(T_i|S_j)\,p(S_j) = p(T_i, S_j)$$

19. $$p(S_j) = p(S_j|T_1)\,p(T_1) + p(S_j|T_2)\,p(T_2)$$

$$\text{since } S = \bar{T}S \cup TS$$

20. $$p(S_j) = p(\bar{T}S) + p(TS);\quad p(S) = p(\bar{T})p(S|\bar{T}) + p(T)\,p(S|T)$$

Equation (15) is the information-theoretical criterion; it is essentially an expression for a decision function which to be minimized. More precisely, the set of sensor-data is partitioned into S_1 and S_2 in such a way as to minimize the conditional

information expression, into which a cost-function has been inserted.

Experimentally this means that a decision can be rendered in accordance with a preassigned decision-criterion say, the conditional entropy criterion, whether a target T belongs to T_1 or to T_2, if it is possible to classify the sensor-data received into dichotomous sets S_1 and S_2. Then the decision is made by minimizing an expression for the conditional entropy, of the type shown in equation (15). $S = \{S_{11}, S_{12}, \ldots, S_{1m}; S_{21}, S_{22}, \ldots, S_{2n}\}$

The aforementioned expression can be made more specific by inserting values for the apriori and aposteriori probabilities, as given below:

21.
$$
\begin{aligned}
p_1 &= p(S_1) \\
q_1 &= 1 - p_1 = p(S_2) \\
p_{11} &= p(T_1/S_1) \\
q_{11} &= 1 - p_{11} = p(T_2/S_1) \\
p_{12} &= p(T_1/S_2) \\
q_{12} &= 1 - p_{12} = p(T_2/S_2)
\end{aligned}
$$

5. <u>RISK FUNCTIONS</u>:

In all decisions one incurs a risk in making a choice; for instance the risk involved that the target T belongs to T_2 when in fact it belongs to T_1, or vice versa.

Let $C_{2,1}$ be the cost of choosing $T \in T_2$ when in reality $T_2 \in T_1$. Similarly, for $C_{1,2}$.

Let $R_1 = C_{2,1}/C_{1,2}$

Now it is assumed that $R_1 > 1$ for the cost ratio to favor $T \in T_1$; likewise, $R_2 > 1$ denotes the fact that the cost ratio $C_{1,2}/C_{2,1}$ favors deciding that $T \in T_2$.

If R_1 is chosen then all the occurrences of p_2, the apriori probability that S belongs to S_2 are multiplied by R_1 in the expression for $H(T|S)$ and the conditional decision criterion is minimized;

$H(T/S)_{R_1}$ is minimized. Similarly, this can be done for $H(T/S)_{R_2}$, if R_2 is chosen; (all occurrences of g_1 are multiplied by R_2 *), should R_2 be greater than R_1).

To sum up: We have a decision rule that enables us to partition the sensor-event space S into two dichotomous sets S_1 and S_2 in such a way that if an observed sensor-event, say S^* belongs to S_j; $(j = 1, 2)$, then T is selected to belong to T_j; $(j = 1, 2)$:

$$(S^* \in S_j) \rightarrow (T \in T_j); (j = 1, 2)$$

The sensor-event-space S is partitioned into S_1 and S_2 with a view of minimizing $H(T/S)_{R_1}$ or $H(T/S)_{R_2}$

6. INFORMATION CONTENT OF A SENSOR PARAMETER:

Let us now turn to considering the amount of information contributed by a single signature parameter of a sensor event. By a signature parameter is meant the characteristic of a response, such as the spectral density of a signal, etc, that helps to recognize or identify a target. The amount of information delivered by a particular value s_{ik} of a sensor-event parameter to the recognition of a target class, pattern, or class is given by:

$$21. \quad I(t_j; s_{ik}) = \log\left[\frac{p(t_j \mid s_{ik})}{p(t_j)}\right]$$

where k stands for the k-th parameter value; k=1 to s_k.

$I(t_j; s_{ik})$ is the amount of information provided by the k-th measurement of the parameter of the sensor-event s_{ik} about the target-event t_j (the target class i.e., that a target t_j belongs or does not belong to target class T).

If an event s_{ik} has occurred, the information supplied by this knowledge alone is given:

$$22. \quad I(s_{ik}) = -\log p(s_{ik})$$

The average value of self information is given by:

$$23. \quad I(s_i) = -\sum_{S_i} p(s_{ik}) \log p(s_{ik}) = H(S_i)$$

$I(t_j; s_{ik})$ can be expressed as:

$$I(t_j; s_{ik}) = I(s_{ik}) - I(t_j \mid s_{ik})$$

This shows that $I(t_j ; S_{ik})$ is symmetrical in t_j and S_{ik}; it is called the mutual information between t_j and S_{ik}.

$$24. \quad I(T_j | S_i) = H(T_j | S_i)$$

$I(T_j | S_i)$ is the average value of the conditional self information.

The average information contributed by all the parameters over the entire sample of i_k discrete parameter values and the j target classes is given by:

$$25. \quad I(T_j ; S_i) = \sum_{T_j} \sum_{S_i} p(t_j, S_{ik}) \log \left[\frac{p(t_j | S_{ik})}{p(t_j)} \right]$$

since

$$p(t_j, S_{ik}) = p(t_j) \, p(S_{ik} | t_j)$$

the information content of the signature parameter S_{ik} is:

$$26. \quad I(T_j ; S_i) = \sum_{j=1}^{n_j} \sum_{k=1}^{n_k} p(t_j) \, p(S_{ik} | t_j) \log p(S_{ik} | t_j) - \sum_{k=1}^{n_k} p(S_{ik}) \log p(S_{ik})$$

The latter expression can be expressed in terms of the entropy function:

$$27. \quad I(S_i ; t_j) = - \sum_{j=1}^{n_j} H(S_i | t_j) \, p(t_j) + H(S_i)$$

$p(t_j | S_{ik})$ = the conditional probability that the target t_j belongs to T_j after a particular parameter observation of S_{ik} the sensor events has been made.

$p(S_{ik})$ = the probability that the given parameter has been measured, i.e., that the parameter measurement has been obtained.

$p(B_{jk}|T_j)$ = the probability that the given parameter observation is obtained from a particular target t_j that belongs to class T_j.

$p(T_j)$ = the a priori probability that a particular target t_j is a member of class T_j.

The average amount of information obtained by a set of signature parameters about a target class T_j is the difference between the entropy of the parameter distribution for all the target classes and the entropy of the distribution of the parameter undergoing observation and measurement.

7. THE DECISION RULE:

In this paper we are concerned with making decisions in selecting a target to T_1 or to T_2. This decision is made upon a predetermined criterion that enables us to separate the class of sensor-events into two dichotomous sets S_1 and S_2. The classification procedure is essentially equivalent to the partition of the sensor-event space and subsequently the target space into at least two mutually exclusive subspaces. Points of the sensor space are assigned to each subset with varying degrees of probability. In order to make sure than an arbitrary sensor event is given a unique assignment to one class a threshold magnitude is selected on the basis of both theoretical judgment and experience.

In our present discussions there are only two classes of sensor events and two sets of target events. It is well known that an optimal decision rule for this case is one in which the test statistic is a likelihood ratio; this ratio is compared with a threshold, say, K.

The a priori likelihood of an event to occur or of a hypothesis to be true is the ratio of two probabilities. The a priori likelihood, for instance, of a sensor event to occur is the ratio:

$$\frac{p(s)}{p(\bar{s})}$$

where $p(\bar{s})$ is the probability that the sensor event S does not occur and the a priori likelihood of a sensor event to belong to class S_1 is the ratio:

$$\frac{p(s_1)}{p(S_2)}$$

We shall as the likelihood ratio the following expression:

$$L(S) = p_1/q_1$$

where p_1 and q_1 are defined in equation 21.

If $L(S) > K$, S is selected to belong to Set S_1

If $L(S) < K$, S is selected to belong to S_2

For a given sensor event S, that class is selected for its set in accordance with the following expression:

$$K = \frac{p_1 C_{2,1}}{q_1 C_{1,2}}$$

This expression is contrived to minimize the average cost. This is the threshold chosen for the likelihood ratio, where $C_{2,1}$ and $C_{1,2}$ are the costs of misclassification as shown in the previous discussions.

GROUND COMBAT COMMUNICATIONS SIMULATION MODEL*

Mr. James W. Virden - Technical Program Director
Robert H. Parke - Specialist 7 - Mathematics-Statistics Assistant

U.S. Army Combat Development Command
Communications-Electronics Agency
Concepts Division

A. Background of Development

A Combat Development Communication Electronic (CD/C-E) Study normally concerns all echelons and elements of command in one or more future time frames (10-20-30 years in the future), therefore, a wide variety of viewpoints must be considered in the preparation of such a study. As the study is developed, the author must figuratively project himself, step-by-step into the following positions:

Platoon: Leader, Member

Commander and his principal staff officers: Company, Battalion, or Brigade

Commanding General and his principal staff officers: Division, Corps, Army, Theater Army

Also of primary concern in a CD/C-E Study are the viewpoints of the Signal or Communications Officer, Signal NCO, and Communications Specialists, into whose roles the author must project himself in the following types of commands:

Combat

Combat Support

Combat Service Support

In addition to placing himself in the above positions the author should further project himself into future time frames and depict the wide variety of environments (geographical, social, political, economic, and technological) expected to be in existence at those times and places.

---

*The opinions expressed in this paper are solely those of its authors, and do not necessarily reflect the opinions or policies of the Agency or Command.

It is obvious that no one individual, or a small group of individuals assembled, can have had experience in all the positions listed above, or even in a majority of the positions. Further, no one has had experience in the future. To provide a semblance of the talent required for such studies, men with years of experience are required to perform much research and study before they can develop a study with any degree of validity. However, if much of the position experience could be placed on a computer, and specialists of various technologies furnish future projections for the computer, and then these experiences and projections correlated and played in actions, more valid studies could be produced with much less learning research, and mental correlation required of the authors. It was this thought of almost ten years ago that generated a contract to General Analysis Corporation to study the feasibility of simulating communications ideas or concepts on a computer. The results of the study indicated that simulation of communications to determine the effects of communications on combat would be feasible, and a subsequent contract was awarded to General Analysis Corporation (later consolidated with CEIR, Inc.) to cover a period of 5 years for development of the model which would be able to simulate the communications of a Type Field Army in combat. Due to a variety of conditions, not the least of which was the state-of-the-art in computers at the time, this complete objective has not yet been met. It was determined to start in the combat area, mainly within the Division, and to work back from there through Corps and Army for the development of the model. Using the IBM 709, a computer in Fort Huachuca, it finally developed that to get sufficient detail for the Division model it was required that computer time would run 4 to 8 times that of actual combat. At the end of the 5 year contract, the Division model was completed to run on the IBM 709. Early in 1965 another contract was awarded to the Philco Corporation to:

Review and update the logic of the model.

Rework the model for the IBM 7090 which is now installed at Fort Huachuca.

Validate the model.

Design a methodology for automation of inputs.

The first two tasks have been completed and work is in progress on the last two. Runs made to date indicate the model is valid. However, the task is not yet complete. The completed tasks have developed the model to the point where it now can be run on the IBM 7090 in a computer time/combat time ratio of 1:1.

B. Summary Description of the Simulator

The Ground Combat Communications Simulation model is a free-running, completely inclosed mathematical and logical model of ground combat between two forces up to Division in size in which communications and information flow is specifically considered. The model permits a detailed observation of communications events in a realistic combat environment, and provides a means of measuring the relative merits of competing communications systems or concepts in that same combat environment. The model augments the usual methods available to the communications system designer by providing:

1. More precise computation of the effects of combat phenomena on the communications system, its organization and doctrine.

2. Information to the designer so he can more explicitly determine the "bottle-necks" or problems likely to occur within a communications system which may be caused by the characteristics of the system, its organization, its doctrine, and the effects of combat.

3. Information to the system designer so he can better determine the comparable effects of the communications systems on the combat in terms of time required to reach an objective, attrition of enemy forces, and attrition of friendly forces.

4. Reports of actions printed out in exhaustive detail, or directly available in computer language for statistical analysis by computer techniques.

5. Consideration of complex interrelated details such as intelligence data flow and forward observers.

6. Easy expansion to include models of phenomena desired to be studied in detail, but which are too complex for manual or mental processes.

7. Ability to quickly and readily repeat games with small changes of initial conditions or parameters.

8. Ability to quickly and readily play variations in the characteristics of equipment and systems not yet in the design or production phase.

C. Significance of the Model

The Ground Combat Communications Simulation model is a significant breakthrough in the development of analytical tools and techniques to study communications and actions in the tactical battlefield. Although

simulations and war games presently exist which can be used to evaluate situations in a combat environment, they generally do not offer the following capabilities which this model offers:

1. Explicit and detailed play of communications and message flow in a dynamic environment - Many simulations have been constructed which play message flow, but only on the basis of statistically generated or predefined message load.

2. Completely inclosed, free-running system - Many simulations and games require human intervention or some other manual manipulation. This makes repeatability of experiments and evaluations extremely difficult, if not impossible.

3. Explicit and detailed play of intelligence and decision interaction - Although a few simulations have played intelligence actions in the tactical battlefield, they generally required use of human judgment since rules of play were not completely explicit or detailed.

4. Wide flexibility of use - Through use of executive control, model compartmentation and ability to change elements of the data base including both initial conditions and parameters. Most simulation and games have only a single orientation and can not be easily modified.

5. Comprehensive modeling of the functions of fire power and maneuver, including the decision and communications process relating the two functions.

In the hands of competent and skilled analysts, the model can prove to be a sophisticated and significant advance over any existing analytical tools for evaluating communications systems, concepts, equipment, operations and doctrine in a combat environment, and the comparable effects between two or more communications systems on combat.

D. Simulator Design

The simulator is structurally an extensive, modularized computer program operating on a large store of data, run on an IBM 7090 computer using an IBM 1301 disk storage unit. Tactical actions of organizations in conflict are determined by logic and input data which specify the Table of Organization and equipment, organization for combat, standard operating procedures, tactical and other doctrine, mission, terrain, environment, and the communication system. Any of these data may be changed by the analyst to study the ability of the communications system under analysis to support a wide range of tactical actions within the limitation of a realizable action.

The simulator proper is a large collection of logical rules and mathematical models for leading two opposing military forces through an exercise in much the same way that Red and Blue teams lead opposing forces over a map in a traditional map exercise. The rules and models of the simulator correspond to the rules and decision processes followed by the players and umpires of a war game.

The following tactical actions are played in detail:

1. Close combat fires including attrition, splitting of fires from armor, infantry, and mechanized elements, and coordination between adjacent elements,

2. Artillery fires including target selection and allocation of both direct and general support batteries,

3. Movement of front line units toward objectives with rear line units adjusting position appropriately,

4. Maneuver:

   a. Offense, to include advance to contact, penetration, envelopment, turning movement, exploitation and reconnaissance in force,

   b. Defense, to include mobile defense, area defense, and combinations thereof,

   c. Retrograde, to include delaying actions and withdrawal,

   d. Commitment and decommitment of Companies, commitment of Battalions and Brigades, and allocation of artillery and other support from reserves,

   e. Selection of lines of departure, character of route and objective,

   f. Movement and deployment of outposts, special units, rear echelon elements and reserves.

5. Communications of all types:

   a. Specifically, tactically essential messages whose individual delivery affects the course of combat,

   b. Generally, tactical and logistical messages whose performance is eventually applied to effectiveness of front line units.

   c. Complete representation of the physical capabilities of communications systems in a combat environment.

6. Combat surveillance, target acquisition and intelligence.

7. Route selection as a function of terrain and intelligence.

The Simulator is an extremely flexible tool in two ways:

First, the user can apply it to any initial conditions within certain broad limitations of the overall capacity of the Simulator. Initial conditions refer to the type of military organization, the number and kinds of units in it, the weapons and other hardware items with which the units are equipped, the terrain on which the exercise will take place, the objectives of the opposing forces, and so on. Of course the Simulator will not operate without these initial conditions, and every application of it will require the specification of these initial conditions. The specification must be in much the same form as would be made for a map exercise.

Second, the extremely large number of parameters which occur throughout the rules and models of the Simulator. The parameters are simply all the numerical values which occur in the rules. Thus, one rule may state that, with no opposition, a certain kind of tank will move 15 miles per hour over a certain type of terrain. The number 15 is thus a parameter. The user can change the parameter to any number he likes and thus explore the implications of differing tank mobility. There are literally thousands of these parameters in the Simulator, corresponding to such effects as artillery range, damage effect of artillery barrages, relative fire power of different weapons, reliability of communications equipment, signal-to-noise ratio as a function of range, etc. The ability of the user to adjust these parameters to his particular purposes makes the Simulator an extremely powerful and versatile analytical tool.

The Simulator is designed primarily as a test environment for the analysis of communications systems performance. Thus, many elements of tactics such as the options available to the command/control decision programs are limited as the model is now designed and are not sufficient for advanced tactical studies. However, the basic programs which (1) compute fire, attrition, movemont rates, artillery damage, suppression and targeting, acquisition and dissemination of intelligence, and which (2) control communications, are considerably more sophisticated than played in other simulations and games. For example, the following actions are cycled in 1 minute, 5 minute, and 15 minute cycles as indicated:

ONE MINUTE CYCLE

Attrition and Move Rate and Movement of Front Line Units

Current attrition and move rates are computed for front line units. The front line units are moved along assigned routes toward their objectives at the current move rate. Status reports are sent to the Battalion.

Coordination of Fires and Intelligence

Adjacent front-line units exchange fire-coordination messages. Front-line units and forward observers acquire intelligence and transmit intelligence messages to Battalion headquarters and fire direction centers respectively.

Direct Support Artillery Fires

Direct support artillery fire: performs target analysis; generates target lists; allocates batteries; fires missions; requests reinforcing fire from the Division Artillery as required.

General Support Artillery Fires

Division Artillery: performs target analysis; allocates missions to general support Battalions; Battalions perform target analysis; allocate batteries; fires missions.

Ground Combat Fires - Artillery Damage Assessment

Computes damage on targets from artillery fires, acquires targets, and computes the suppressive effect of artillery fires. Selects optimum weapon-target match for closed combat weapons. Computes the amount and type of fire received by each unit in contact. Computes force ratios.

Implement Battalion Decisions

If the Battalion commander has decided to commit or decommit a unit: change unit commit status; move the unit to its new location; generate status reports to Brigade.

Process Messages

Process all messages over the communications system. Processing entails: message center procedures; encryption and decryption of messages; route selection and switching; implementing "busy" doctrine.

FIVE MINUTE CYCLE

Generate Administrative and Logistic Messages

This message load is generated through the Division area. The tactical effect of communications system performance is determined.

Units in Contact

From the list of major elements in contact, opposing pairs of units that have the potential for contact in the next five minutes are determined and listed.

Attrition and Move Rates - Pattern Movements

Current attrition and move rates are computed for all units not in a committed status. Movement orders are generated to units in patterns as required. Units in movement patterns that have received movement orders begin or continue movement to new locations.

Communications System Status

Determine current status of all circuits. Status is: operable; degraded, or inoperable. Status is a function of: distance; damage; reliability, radio frequency interference.

Battalion Commitment or Decommitment Decisions

Battalion commanders evaluate intelligence and the status of front-line units to determine the necessity for committing a reserve Company or for decommitting a front-line unit.

Implement Brigade Commit Decisions

If the Brigade Commander has decided to commit: Commit orders are sent to the Companies; Battalion commit status is altered; a route to the pre-selected front-line position is chosen; status reports to the Division are generated.

Implement Division Commit Decisions

If the Division commanding general has decided to commit: a commit order is sent to the Battalion commander after staffing delay; the Brigade commitment status is altered; and routes are selected for the movement of Battalions into their front-line positions.

FIFTEEN MINUTE CYCLE

<u>Major Elements in Contact</u>

Determine and list opposing groups of units that will come into contact with one another in the next 15 minutes; the basis for the range of surveillance devices; and the basic move rate.

<u>Intelligence Acquisition</u>

The reconnaissance units having a long-range surveillance capability acquire intelligence about enemy units.

<u>Intelligence Reports</u>

Reconnaissance units transmit intelligence reports to the Division headquarters, Division artillery, and Brigade headquarters.

<u>Dissemination of Intelligence</u>

Intelligence reports are exchanged by all headquarter units.

<u>General Outpost Line Decisions</u>

Status of the general outpost line is checked to determine whether or not withdrawal is in order. If so, the general outpost line units begin moving to pass through blocking units in the initial delaying position.

<u>Division Commit Decisions</u>

The Division commanding general evaluates intelligence and status of the main effort Brigade to determine the necessity of commitment of the reserve Brigade. If necessary for commitment, the commit order is generated.

<u>Brigade Commit Decisions</u>

Brigade commanders evaluate intelligence and status of commited Battalions to determine the necessity for commitment of a reserve Battalion. If the commitment is necessary, a commitment order is generated.

<u>Determine Terrain Values</u>

This assign terrain classification values to all units based upon their map location.

The Simulator is currently programmed in FORTRAN II, Version 3 programming language for operation on an IBM 7090/1301 computer system or equivalent. The simulation as now constituted operates in approximately real time; that is, one hour of game time is equivalent to one hour of actual combat.

E. Concept of Operation

The model is a mathematical model which operates using two major elements: a logical structure, and an operational data base comprised of tactical data and communications data (see Figure Nr 1).

1. The Logical Structure - consists of three basic models (see Figure Nr. 2).

   The Tactical Model - which simulates for each unit the operation of fire, movement, attrition, intelligence, decisions, and the need for messages and their generation.

   The Communications System Model - which simulates the actual communication system and evaluates the current status of its components. (i.e. Equipment damage and failure; operability of link because of movement of units; radio range; wire failure; radio interference).

   The Message Traffic Processing Model - which simulates the actual flow of messages as they undergo communications center processing, coding, handling, route selection, route availability and switching, transmission, and delivery delays.

2. The Operational Data Base - consists of two major groups:

   Tactical Data - represents the "initial conditions" or scenario and includes information on:

   Terrain
   Tactical organization and disposition
   Tactical situation
   Missions and objectives
   Other physical factors

   Communication Data - includes information on:

   Communications equipment and systems
   Communications operation, including nets
   Communications doctrine, including operating procedures

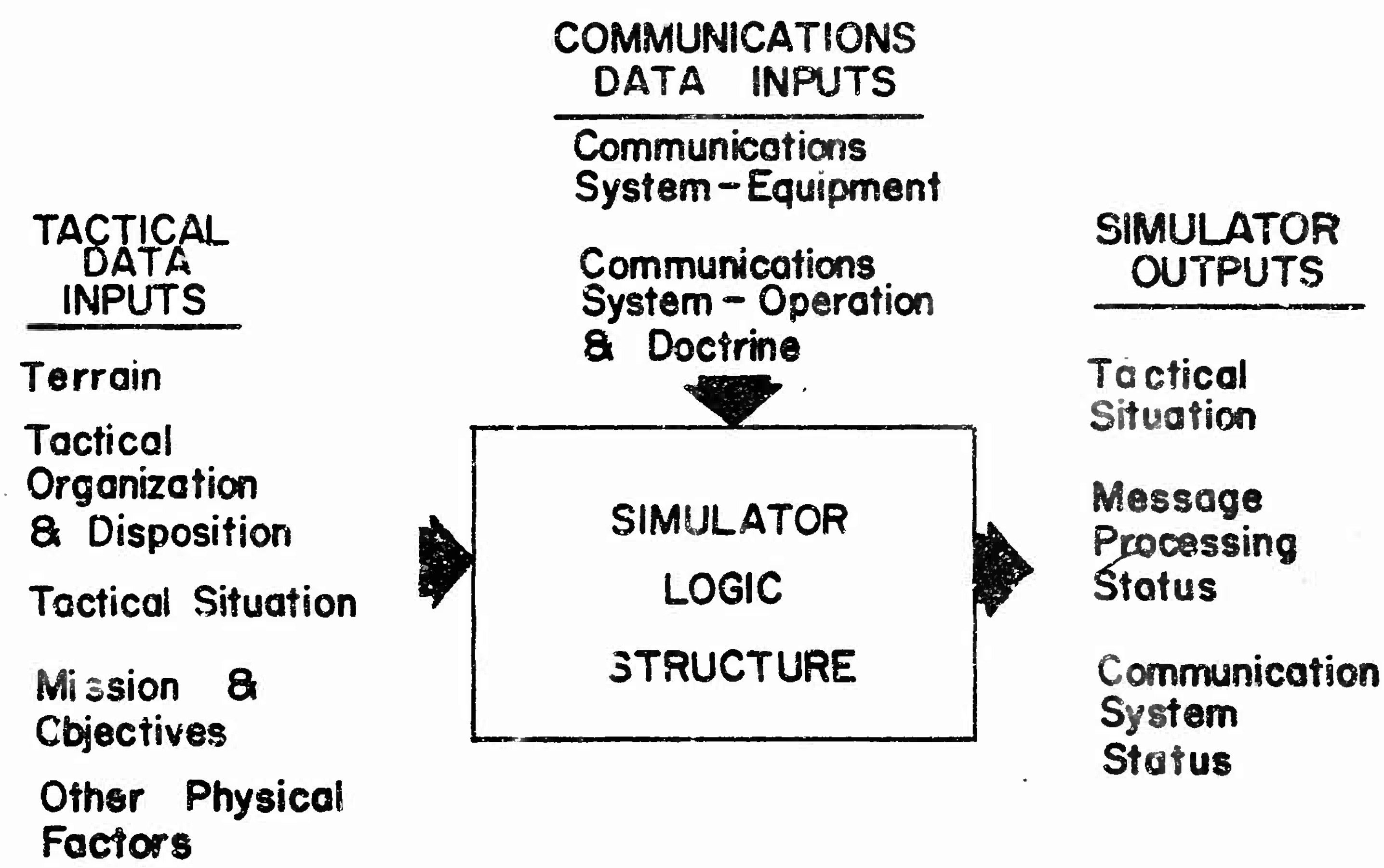


Figure 1. Concept of Simulator Use

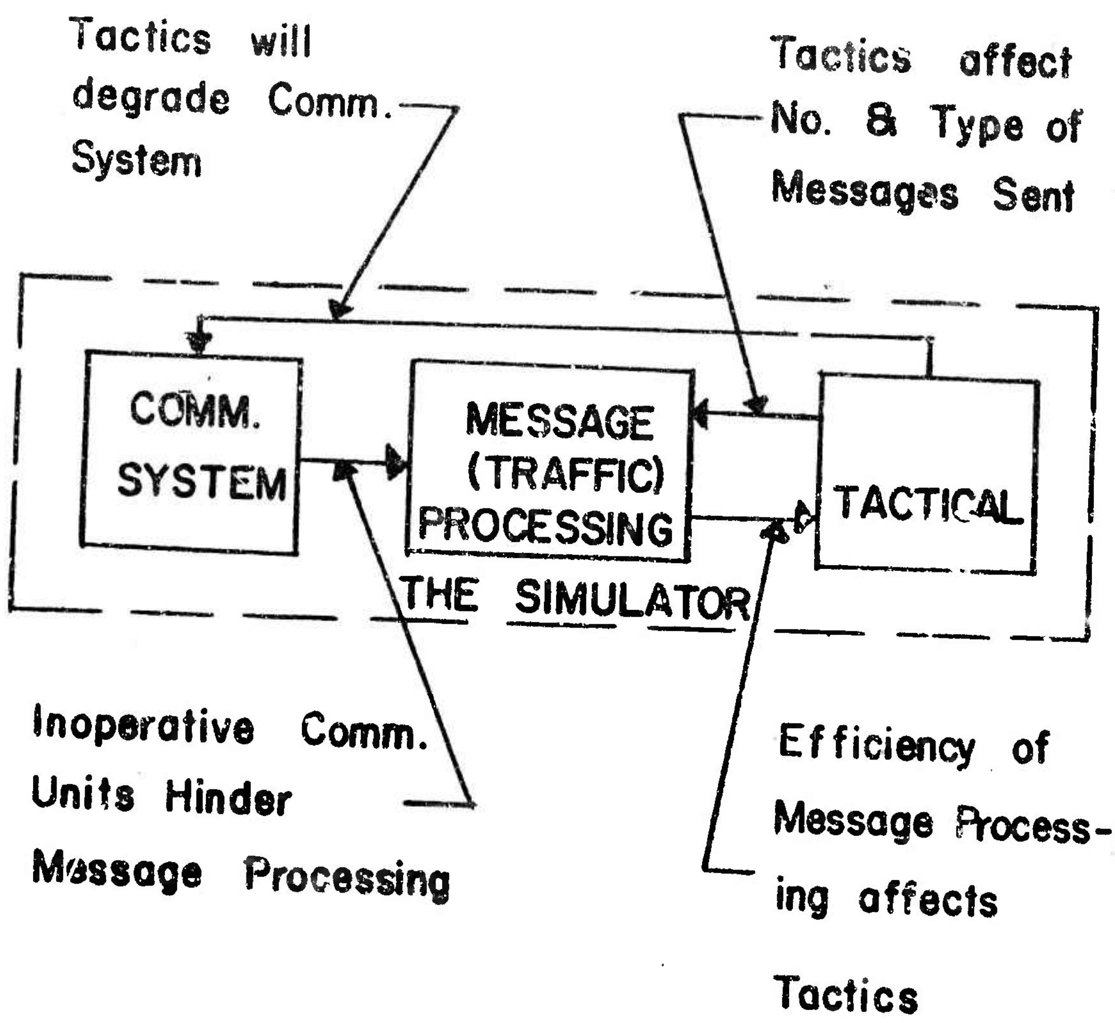


Figure 2. Simulator Organization and Interaction

To use the Simulator, the analyst first inputs the programs and data describing the Simulator's logic structure to the computer. Next, the analyst inserts the operational data base to be used in the evaluation, consisting of the tactical data inputs and the communications data inputs.

The tactical data inputs describes the terrain over which the tactical action will take place, and describes the tactical organization and disposition, together with the concept of operations for both the friendly and enemy forces, missions and objectives.

The communications data inputs are descriptions of the communications system equipment, operations and doctrine to be tested.

Using these initial conditions, a simulation run is made to evaluate alternative communications concepts. Appropriate changes are made to the communications data and new runs are made. To evaluate the effects of tactical organization, disposition, mission, etc, on a given communication concept, appropriate changes are made to the tactical data and additional runs are made.

This will produce three types of output information:

1. Tactical Situation Reports describing the flow of the tactical action in terms of unit location, attrition levels, and amounts of fire delivered or received.

2. System Status Reports indicating which circuits within the communication system are inoperable due to movement of the various units, signal equipment damage or failure, effects of range, and the effects of unintentional radio interference.

3. Message Processing Reports describing the processing steps that have been accomplished for each message being transmitted over the communication system, together with the time the message failed to complete processing, or the time it reached its destination. These reports constitute complete histories for each message being processed over the communication system.

F. Communication System Evaluation Methodology

The method of using the Simulator in evaluating communication systems is briefly described. The tactical configuration and the initial conditions of the concept of operation for the tactical action are determined.

Communications System "A" is then designed and played in the computer in the context of the tactical situation. Outputs are then evaluated subjectively. Normally this evaluation can be accomplished from two viewpoints, depending on the desire of the evaluator:

One method would be to determine from the tactical output, the time required for the friendly force to achieve its objectives, the amount of attrition dealt the enemy, and the amount of attrition incurred by the friendly forces.

The other method would be to evaluate the communication output and determine where "bottle-necks", or problem areas, might be occurring.

Competing communication system "B" is then designed and evaluated on the computer, using the same tactical problem. Based on the results of these comparative computer runs, one can analyze the data and determine the relative merits of the two competing systems. It must be remembered that this comparison is not accomplished by the Simulator itself, but by subjective evaluations made by competent communication systems designers and/or tactical officers, based on data outputs produced by the Simulator.

From the outputs of the computer runs, evaluations can be made of the effects certain characteristics of communications systems have on given tactics in combat, or the effects that different tactics have on a given communication concept.

G. Current Scenario and Use

Several scenarios have been constructed during the process of development, and used for test purposes. The largest one, and the one now in use, describes initial conditions involving an Armored Division against a reinforced Mechanized Brigade where the units range from Company to Division in size. This Division level scenario is presently available for experimental runs, and is the one which is being used in the current contractual effort. This scenario is called "Goldleaf".

The "Goldleaf" data base presently simulates the action of an Armored Division attacking a reinforced Mechanized Brigade, with both sides organized under current tables of organization and equipment. The scene is a varied piece of terrain in the Fulda Gap area of Germany. The attacking Blue Corps of the NATO forces, of which the Armored Division is a part, has the mission of seizing the eastern exits of the Thuringian Mountains in the vicinity of Eisenach. The Corps has planned this mission in three phases:

Phase 1 - Establishing a bridgehead on the Lahn river.

Phase 2 - Seizing crossings over the Fulda river.

Phase 3 - Capturing the Eisenach Mountain exits.

The game is initiated with Phase 1 completed and the Corps reserve, the pertinent Blue Armored Division, committed. The committed Division then advances at full speed, in an effort to seize the Alsfield area

with its approaches to the Fulda river. On the opposing side, a Red Corps has been fighting a delaying action with strongly reinforced Armored Cavalry units. Upon losing the Lahn river position, Red Corps decides to defend the Alsfield area with a reserve Mechanized Infantry Division, thus hoping to hold the area until the arrival of reinforcements.

The action on the Blue side will cause deployment of all Battalion task forces, commitment of Brigade reserves, and probably, commitment of the Division reserve. On the Red side, the reinforced Armored Cavalry regiment manning the Division general outpost line will be driven in, all forward defense units of the strong Red Brigade will become engaged, the Brigade reserves will be committed, and possibly the Division reserve will become committed. The Blue Armored Division will have to cover about 33 kilometers to reach its objective, and the Red Division will be controlling elements in contact over a depth of about 25 kilometers. Under usual circumstances it will take about eight hours or less for the Blue Armored Division to reach their objectives. If they have not reached their objectives by eight hours, it is assumed that the Red Forces reserves will have arrived in sufficient size, quantity, and quality to preclude the Blue Forces from achieving their objectives. Approximately 250 units of organization are played in the attack. The communication system of either the attacker or defender or both can be evaluated.

Currently, the validity test is using this scenario and will permit the evaluation of the use of a Radio Central AN/USC-3 in specified organizations in lieu of radio-wire integration units. The results of the evaluation will then be compared to the results of two recent large scale field maneuvers which used the Radio Central AN/USC-3 in a similar manner.

H. Summary

In essence, the Ground Combat Communications Simulation Model, as it stands today, is the start of automating part of the work required in developing Combat Development Communications Studies. It is that portion of work which needs exorbitant experience and training of authors, and which also forces the author to imagine himself in the future of 10-20-30 years - a talent few have. By modifying emphasis from communications to other areas of study, these same techniques can be applied to other complex Army problem areas.

OBJECTIVE CRITERIA

FOR

ELIMINATING LOWER ATMOSPHERIC EFFECTS

FROM

FIELD TESTS OF ARMY COMMUNICATIONS-ELECTRONICS SYSTEMS

BY

Mr. Kenneth M. Barnett

ATMOSPHERIC SCIENCES LABORATORY
U. S. Army Electronics Command

ABSTRACT

Charts give objective decisions as to whether or not lower atmospheric refraction, absorption, scattering, or ducting will introduce significant effects into electromagnetic propagation data gathered as part of systems test in southern Arizona.

If all charts give negative results, then concurrent observations of atmospheric conditions can be eliminated from the test.

If one or more charts give positive results, then the requirement for concurrent observations of atmospheric conditions has been identified.

Objective Criteria

for

Eliminating Lower Atmospheric Effects

from

Field Tests of Army Communications-Electronics Systems

by

Mr. Kenneth M. Barnett

Atmospheric Sciences Laboratory
U. S. Army Electronics Command

INTRODUCTION

The U. S. Army Electronics Proving Ground at Fort Huachuca, Arizona, is one of the biggest single users of the meteorological services that are provided to RDT&E activities throughout the Army by Army Meteorological Teams. The teams are provided by USAERDAA which is also at Fort Huachuca. These met services for USAEPG include meteorological observations taken simultaneously with field tests of radio and microwave communication, aircraft navigation, radar, surveillance and electronic warfare systems. The purpose of these observations is to allow for corrections in electromagnetic propagation data (signal strength, radar position, etc) which have been altered by atmospheric refraction, ducting, scattering or absorption.

It is obviously important that the "correct" meteorological observations be taken. The cost of the met observations themselves and the cost of correcting the propagation data for atmospheric effects can reach several hundred thousand dollars in a year. Of even greater concern is the question "Were the right atmospheric observations made and were the right corrections made to the propagation data?"

For the past three years a small research program has been conducted by USAERDAA to see if an objective procedure could be devised for determining the right kind and the right amount of met observations that should be taken as part of the field tests of Army Communications Electronics Systems (CES) performed by the USAEPG throughout southern Arizona.

This paper represents the first of two steps that are required for a systematic and objective planning of the atmospheric observations needed as part of the data collection phase of any field test of a specific CES. The other step was presented by the author at the 1964 Army Operations Research Symposium in a paper entitled "The Planning of Optimum Meteorological Observations for Army RDT&E Activities". These two papers supplement each other and combine into a total procedure for planning the atmospheric observations required as part of the field test of a CES.

This paper presents the first step that should be taken by an engineer who is planning a CES field test and is confronted with the question "Do I need observations of the atmospheric conditions during the test and if so, what kind and how many observations are needed?" The first step is essentially negative in that it defines the atmospheric effects which can be ignored or eliminated from the field test. If all atmospheric effects could be eliminated in this way, then the field test can be safely planned without any concurrent observations of atmospheric conditions. (This makes a distinction between "observational services" which is part of the data taking and "forecasting services" which are needed for scheduling test periods and for avoiding weather conditions hazardous to men or equipment. This paper does not attempt to provide criteria for such forecasting services.)

If the first step gives a "yes" answer that atmospheric refraction or ducting, for example, will significantly affect some of the propagation data then it is necessary to take step two. A procedure for this second step was reported in the 1964 Army Operations Research Symposium, as noted above. This procedure may or may not be simple but it is straightforward and, if sufficient theoretical and empirical information exists, it can be made objective. The second step is summarized here

1. Define the RDT&E problem influenced by the atmosphere.
2. Express the physical relation between the problem and the significant atmospheric factors.
3. Determine the allowable error in atmospheric factors that corresponds to the accuracy tolerance for the RDT&E problem.
4. Find the most economical meteorological instrumentation, site spacing, and schedule of observations that will measure the atmospheric factors within the allowable error.

This paper is concerned with step one, which objectively defines those field tests of CES in which the atmosphere will have no significant effect on the propagation data and hence can eliminate the need for atmospheric data as part of the field test.

This paper has incorporated empirical data which make step one applicable to CES tests in southern Arizona. It is assumed that by using different empirical data similar charts could be made for a different area if required.

DISCUSSION

If one considers an electromagnetic wave progressing through a gas, the gas can do several things to the wave; the gas can refract the wave and as an extreme case duct or trap the wave; it can scatter the wave; and it can attenuate the wave by absorption.

This list is not complete but it is considered a reasonable list of the most important effects that the lower atmosphere will have on the short distance propagation paths of the CES tested in southern Arizona. This obviously rules out ionospheric effects. In the electromagnetic spectrum only wavelengths longer than visible light (i.e. infra-red, microwave and radio) are considered.

Reflection effects of the atmosphere were not included since it was assumed that this can be considered as a limiting case of refraction. This is receiving further attention.

A series of six charts have now been constructed which display these different atmospheric effects as follows:

Chart 1. Atmospheric Scattering. This shows the propagation frequency, path distance and antenna heights for which a contribution in the received signal strength of 6 decibels or more because of scattering from the lower atmosphere (or "tropospheric" scattering) can be expected.

Chart 2. Atmospheric Absorption. This shows the frequencies at which absorption by any one of ten gases in the atmosphere can be expected.

Chart 3. Atmospheric Refraction on Radar. This shows the typical radar refraction error in the Fort Huachuca area for ranges and heights from the radar. It is based upon both theoretical calculations and actual observations of radar refraction errors.

Chart 4. Atmospheric Refraction Effects on Line of Sight Distance. This shows how atmospheric conditions will cause the radio horizon to vary. This also requires some information about the frequency of occurrence of atmospheric refractivity gradients as a function of season and time of day. This latter information could be considered specialized climatological information.

Chart 5. Atmospheric Ducting. This is a purely theoretical prediction of the critical frequencies for ducting under certain atmospheric conditions. This also requires some

information about the frequency of occurrence of atmospheric refractivity gradients.

Chart 6. Atmospheric Refractivity Climatology. This is a purely empirical chart which shows the probable frequency with which certain atmospheric conditions are likely to occur. This chart is ancillary to Charts 4 and 5.

With these six charts, an engineer planning a CES test in southern Arizona can very quickly obtain an answer to the question "Will the atmosphere significantly influence the propagation data on this test?" Charts 1, 2, and 3 give a "yes" or "no" answer. Charts 4 and 5 used in conjunction with Chart 6 give a percentage probability answer.

If all answers are "no" or of low enough probability, then he can ignore the atmospheric effects. Thus he has objective criteria for eliminating (lower) atmospheric effects from a field test of CES.

If one or more of the answers is "yes" or of high enough probability, then he has defined which atmospheric effects are significant to his test and for which some atmospheric observations must be made as part of the field test. He can then proceed with the procedure in "The Planning of Optimum Meteorological Observations for Army RDT&E Activities" to determine what kind and how many atmospheric observations are needed.

EXPLANATION OF CHARTS

Chart 1. Atmospheric Scattering. This was determined by the use of both theory and empirical data. The received power was calculated for a diffracted field over a spherical earth of 4/3 radius. This was compared with radio propagation data obtained in south central Arizona. When the observed power was more than 6 db greater than the theoretical it was considered that scatter mode was significant. These data were obtained by the U. S. Navy Electronics Laboratory during 1946-1948.

The derivation of this chart is more fully explained in the U. S. Army Technical Report ECOM-0268-1 "Criteria for Determining when Scatter Mode is Dominant" by John B. Smyth (under contract with Smyth Research Associates), November 1965, published by USAERDAA, Fort Huachuca, Arizona.

Example: The variables are frequency, distance and heights of the transmitting and receiving antenna. If a test requires transmissions at $2x10^3$mcs, then for antenna heights of 12 feet or lower, scatter propagation is significant at all distances; for antenna heights of 24 feet, scatter propagation is not significant for distances less than 25 miles: for antenna heights of 48 feet, scatter is not significant for distances less than 30 miles.

Chart 2. Atmospheric Absorption. This summarized an initial survey of literature. This was prepared by 1st Lt Richard Orville during a short annual U.S. Army Reserve duty tour. It is based upon approximately 20 reports from the National Bureau of Standards and from U.S. Air Force contractors. It is hoped that further work can be done on this.

Example: At a transmission frequency of 15-20 Gcs, no significant atmospheric absorption should be expected. At a frequency of approximately 25 Gcs, significant absorption from SO_2, NO_2 and H_2O can be expected.

Chart 3. Atmospheric Refraction Errors for Radars. This chart applies only to radars located at Fort Huachuca. It indicates a "typical" error that can be expected if the radar beam is assumed to go in a straight line. This error is almost entirely (>90%) in the vertical with the uncorrected radar, reporting an aircraft to be higher than it actually is. These actual errors vary continuously with changing atmospheric conditions and in extreme cases, might be different by factors ranging from 1/2 to 2.

This chart was first calculated by a formula developed in report USAERDAA-MET-7-64 "A Review of the Calculation of Radar Refraction Errors" by Barnett, Bomba, Heil and Kirchner, June 1964 and published by USAERDAA, Fort Huachuca, Arizona. The first version of the chart was contained as Figure 2 in report USAERDAA-MET-9-64 "An Objective Procedure for Planning Meteorological Observations Needed to Calculate Radar Refraction Errors" by Barnett June 1964. This chart was revised to make it consistent with observed data contained in Annex C to report ECOM-6004 "A Comparison of Observed and Calculated Radar Refraction Errors" by Barnett and Brown, July 1965. The variability of this error at low elevation angles is contained in report USAERDAA-MET-2-65, January 1965 and Addendum, May 1965 by Carlson.

Example: If an aircraft or drone position during flight must be measured by radar to a certain accuracy, this chart will immediately tell whether or not the radar readings must be corrected for atmospheric refractivity. If the aircraft will go a maximum distance of 60 miles from the radar and will be as low as 10,000 feet msl, then the uncorrected radar will report the aircraft to be about 500 feet too high. If the required precision of the aircraft position is more than 500 feet, then atmospheric refraction can be ignored.

Note: If much finer precision is required, Chart 3A gives some indication of the accuracy to which the corrections to the radar can be calculated with the use of very minimum atmospheric observations, namely refractivity observations at the radar site and monthly mean values at about 20,000 feet msl. This is also reported in technical report ECOM-6004.

Chart 4. Atmospheric Refraction Effects. This is entirely theoretical and applies to the determination of the radio horizon and assumes a spherical earth. Chart 6 must be used in conjunction with this chart. It shows how the radio horizon will vary with lower atmospheric conditions. When the refractive index decreases rapidly with height, the radio wave is bent downward and the distance to the radio horizon becomes greater. This chart shows that theoretically ducting can occur when the atmospheric refractivity decreases vertically at a rate greater than 48N units per 1000 feet.

The derivation of this chart is explained in report ECOM-6010 "The Influence of Atmospheric Refraction on Electromagnetic Propagation" which was edited by 2nd Lt Neil M. Schmitt, November 1965.

Example: The variables are antenna height, distance to radio horizon for a spherical earth, and atmospheric refractivity. If a transmitter and receiver are both at 20 feet and we want to ensure that they are not in a line of sight more than 50% of the time over an extended period of several weeks, we go to Chart 6 and find that for 50% of the time in southern Arizona the vertical gradient of refractivity in the lower 1000 feet is -24N units per 1000 feet or greater (i.e. -20 or -10 units per 1000 feet). Now go to Chart 4 and find the diagonal line for G = -24 and go down the diagonal line to the intersection of antenna height of 20 feet. Then go down the vertical line to the abscissa and find 7.8 miles. The minimum required distance between the transmitter and receiver would then be twice that or 15.6 miles. If the transmitter and receiver had to be closer than that, then line-of-sight transmission could be expected more than 50% of the time.

Note: Chart 6 is a composite for all seasons and all times of day. If the test were for only a certain season or a certain time of day or a certain general weather condition, a more specialized version of Chart 6 would be required.

Chart 5. Atmospheric Ducting. This is a theoretical chart and required three significant assumptions for its derivation. It should be used with Chart 6.

Report ECOM-6009 "The Influence of Atmospheric Ducting on Electromagnetic Propagation" edited by 2nd Lt Neil M. Schmitt, November 1965, explains the derivation and assumptions.

Example: No ducting should occur for vertical gradients smaller than -24N units per 1000 feet. (-10N units per 1000 feet would not produce ducting). Chart 6 shows that this should occur about 50% of the time over a long period.

The depth of the atmospheric layer with this refractivity gradient (H) is also a factor. This layer can be very shallow

(100 feet for example) and produce a duct at frequency of 40 Gcs while the same gradient (-24N/1000 ft) would have to be 1000 feet deep to produce ducting at 100 mcs.

Note: Report ECOM-0268-2 "Refractive Index Gradients Gila Bend-Dateland, Arizona Area" by John B. Smyth (under contract with Smyth Research Associates) November 1965, gives some frequency distributions of the refractivity gradients through different depths of the atmosphere. It can be used to supplement Chart 6.

Chart 6. Atmospheric Refractivity Climatology. This is the compilation of 679 radiosonde, tower and captive balloon observations of refractivity in the lower atmosphere from Fort Huachuca to Yuma, Arizona, with the bulk of the data taken at Gila Bend, Arizona.

Report ECOM-0268-2 supplements this and is based on 160 separate soundings.

The use of this chart was explained in examples to Charts 4 and 5.

Many approximations and assumptions have been made in developing these charts. It is valuable to verify these charts with actual measurements of field strength, radar refraction errors and other values. Some experiments to measure refraction errors have already been completed and experiments are planned in the next few months to measure field strength due to different propagation modes.

CONCLUSIONS

A set of charts are available to tell the engineer who is planning a CES field test whether or not atmospheric refraction, scattering, ducting or absorption will significantly influence the EM propagation data to be gathered as part of the field test.

If all charts indicate that the atmosphere will not significantly influence the test data then no concurrent atmospheric observations will be needed as part of the test. This "eliminates" atmospheric effects from field tests of Army communications-electronics systems tests.

If the charts do indicate that one or more atmospheric effects will significantly influence the test data, then the engineer has defined the kind of atmospheric observations that must be made concurrently with the CES propagation observations. A subsequent method to determine the exact kind of meteorological observations required and how many required was explained at the 1964 Army Operations Research Symposium in a paper entitled "The Planning of Optimum Meteorological Observations for Army RDT&E Activities".

REFERENCES

1. Barnett: "The Planning of Optimum Meteorological Observations for Army RDT&E Activities" Proceedings, U.S. Army Operations Research Symposium (Rock Island, Ill) May 1964.

2. Barnett, Bomba, Heil, Kirchner: "A Review of the Calculation of Radar Refraction Errors" USAERDAA-MET-7-64, June 1964.

3. Barnett: "An Objective Procedure for Planning Meteorological Observations Needed to Calculate Radar Refraction Errors" USAERDAA-MET-9-64, June 1964.

4.5. Carlson: "An Analysis of Atmospheric Variability Pertinent to Low Level Radar Tracking Operations In Arizona" USAERDAA-MET-2-65, Jan 1965, and Addendum, May 1965.

6. Barnett, Brown: "A Comparison of Observed and Calculated Radar Refraction Errors" ECOM-6004, July 1965.

7. Schmitt: "The Influence of Atmospheric Ducting on Electromagnetic Propagation" ECOM-6009, November 1965.

8. Smyth: "Criteria for Determining when Scatter Mode is Dominant" ECOM-0268-1, November 1965.

9. Smyth: "Refractive Index Gradients, Gila Bend-Dateland, Arizona Area" ECOM-0268-2, November 1965.

10. Schmitt: "The Influence of Atmospheric Refraction on Electromagnetic Propagation" ECOM-6010, November 1965.

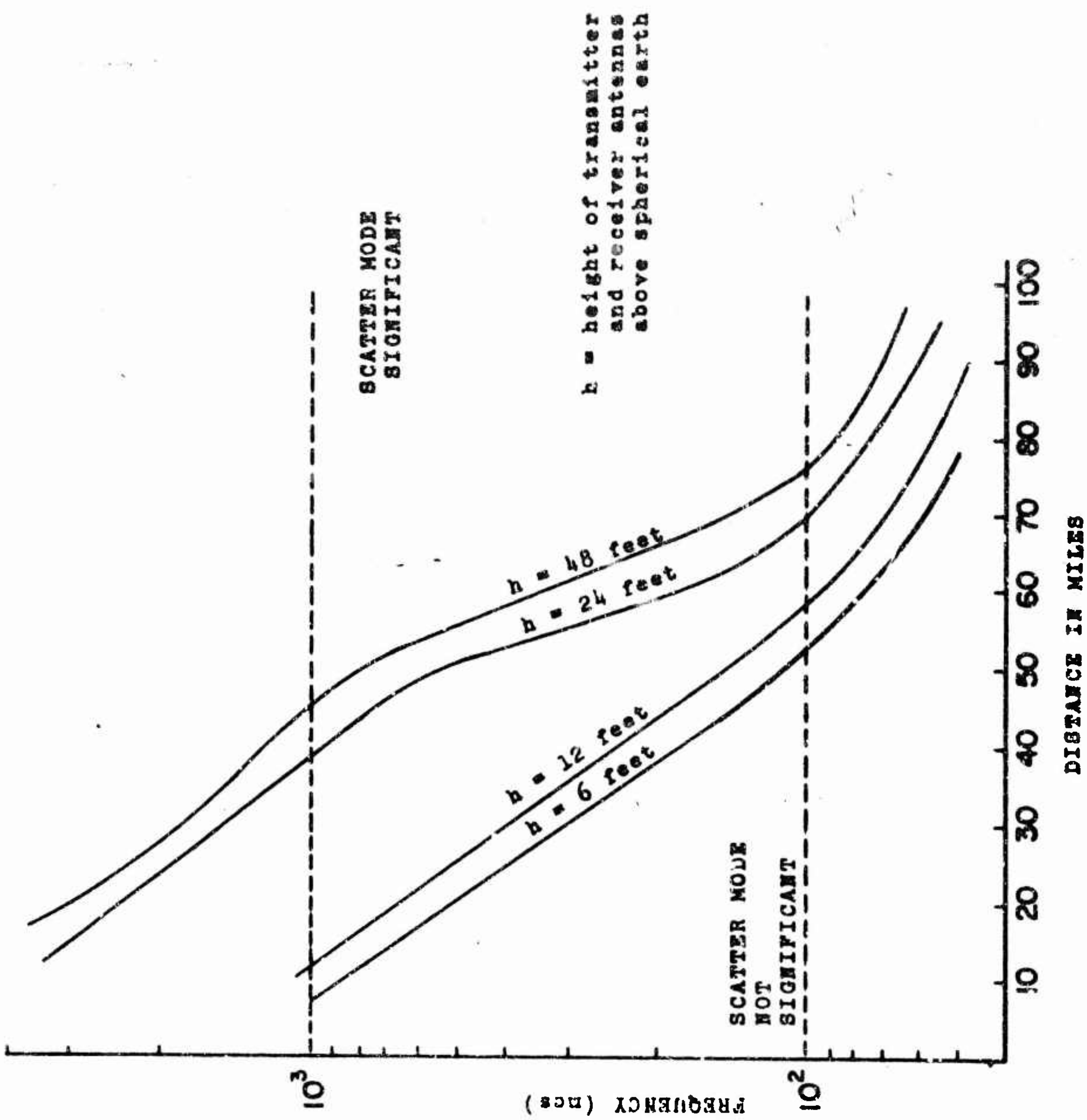


CHART 1
ATMOSPHERIC
SCATTERING

Source: Figure 2
Technical Report
ECOM 0268-1

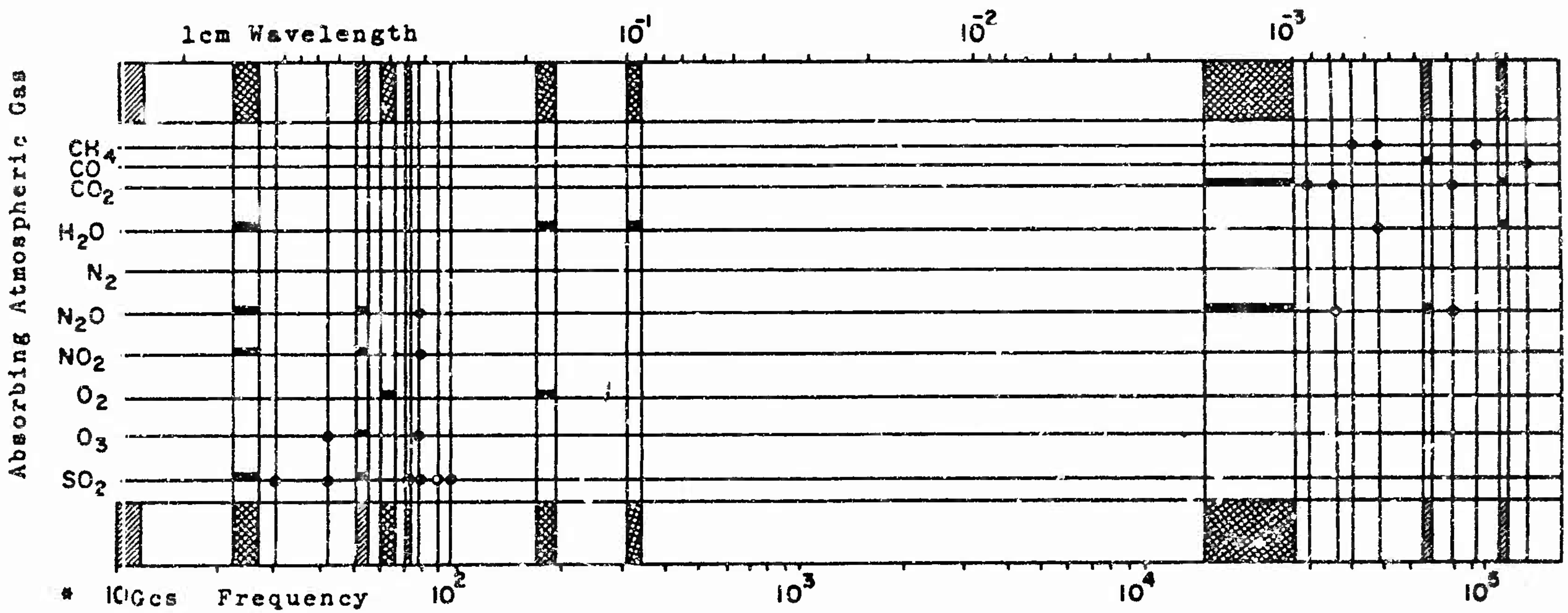


CHART 2. SUMMARY OF ABSORPTION BANDS

Double cross hatching indicates most significant absorption bands.

*Note: Minor amounts of absorption occur between 1 - 10Gcs due to O_2.
Little is known about absorption between 10^3 to 10^4 Gcs.

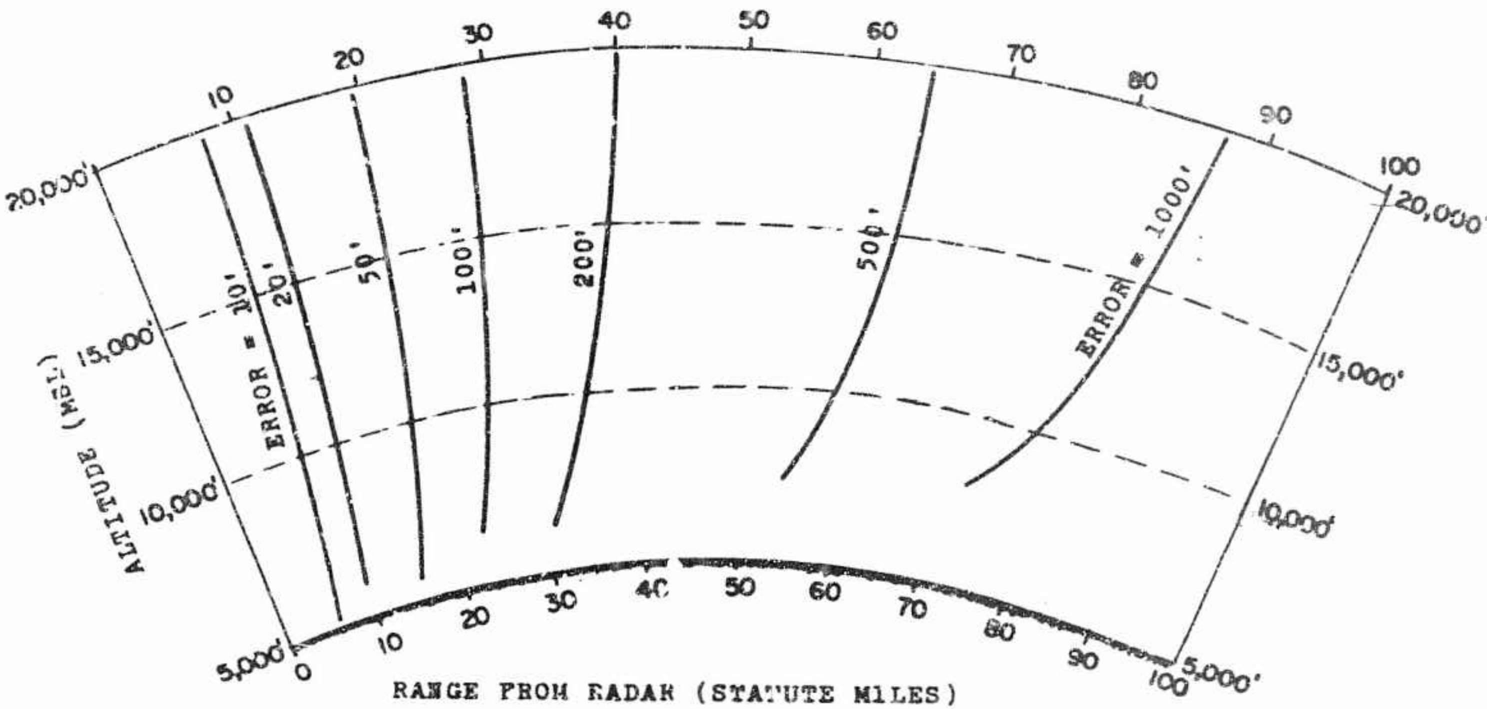
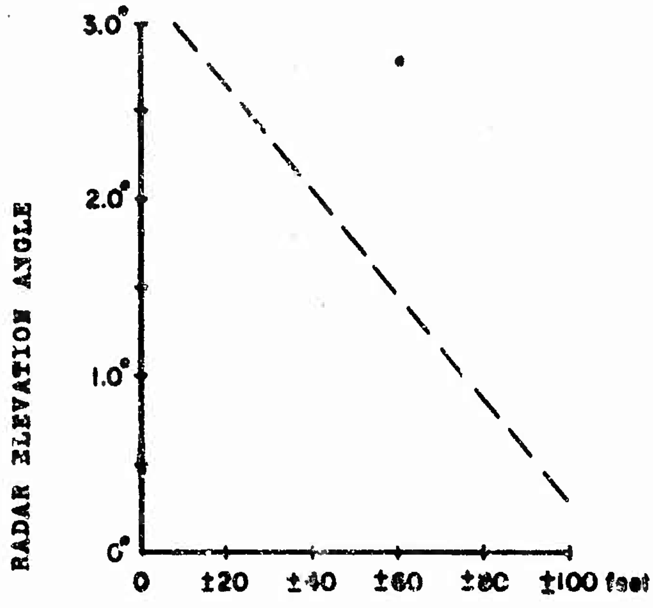


CHART 3. ATMOSPHERIC REFRACTION ERRORS FOR RADARS AT FORT HUACHUCA, ARIZONA

Source: Figure 2 from report USAERDAA-MET-9-64 as corrected by data in Annex C of report USAERDAA-MET-4-65

CHART 3A. PRECISION TO WHICH RADAR REFRACTION ERRORS CAN BE CALCULATED BY USE OF SURFACE REFRACTIVITY OBSERVATIONS



RMS VALUE OF PRECISION TO WHICH RADAR REFRACTION ERROR CAN BE CALCULATED. RANGE 40 MILES

This data is limited to data taken along an azimuth of 321° from the FPS-16 radar at Fort Huachuca during daylight. The method used is a computer program RA-112 and is based on an atmospheric refractivity condition which decreases exponentially from the surface observed value of the refractive index.

See report MET-4-65

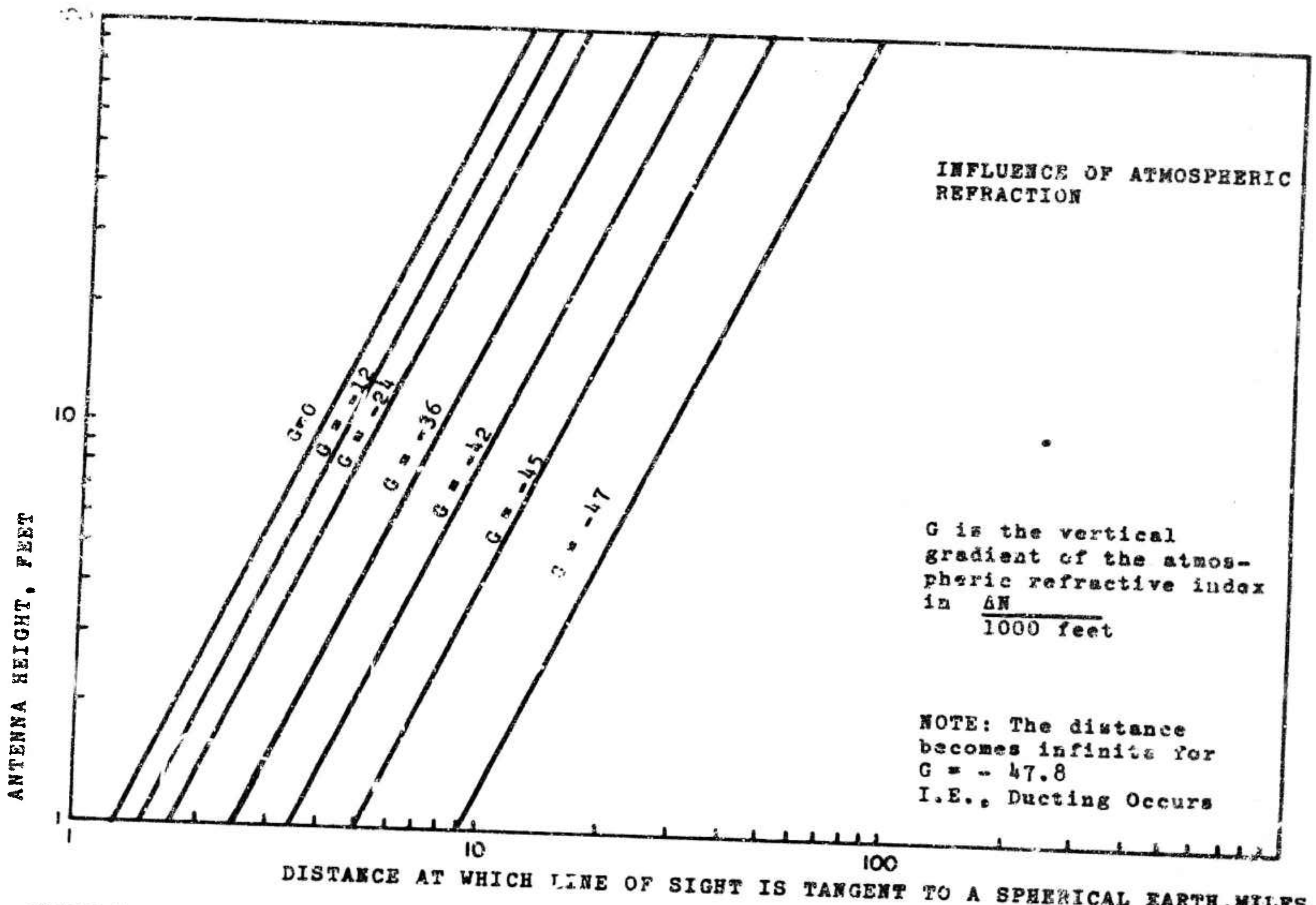


CHART 4. ATMOSPHERIC REFRACTION EFFECTS ON LINE OF SIGHT DISTANCES

Source: ECOM-6010, Figure 1.

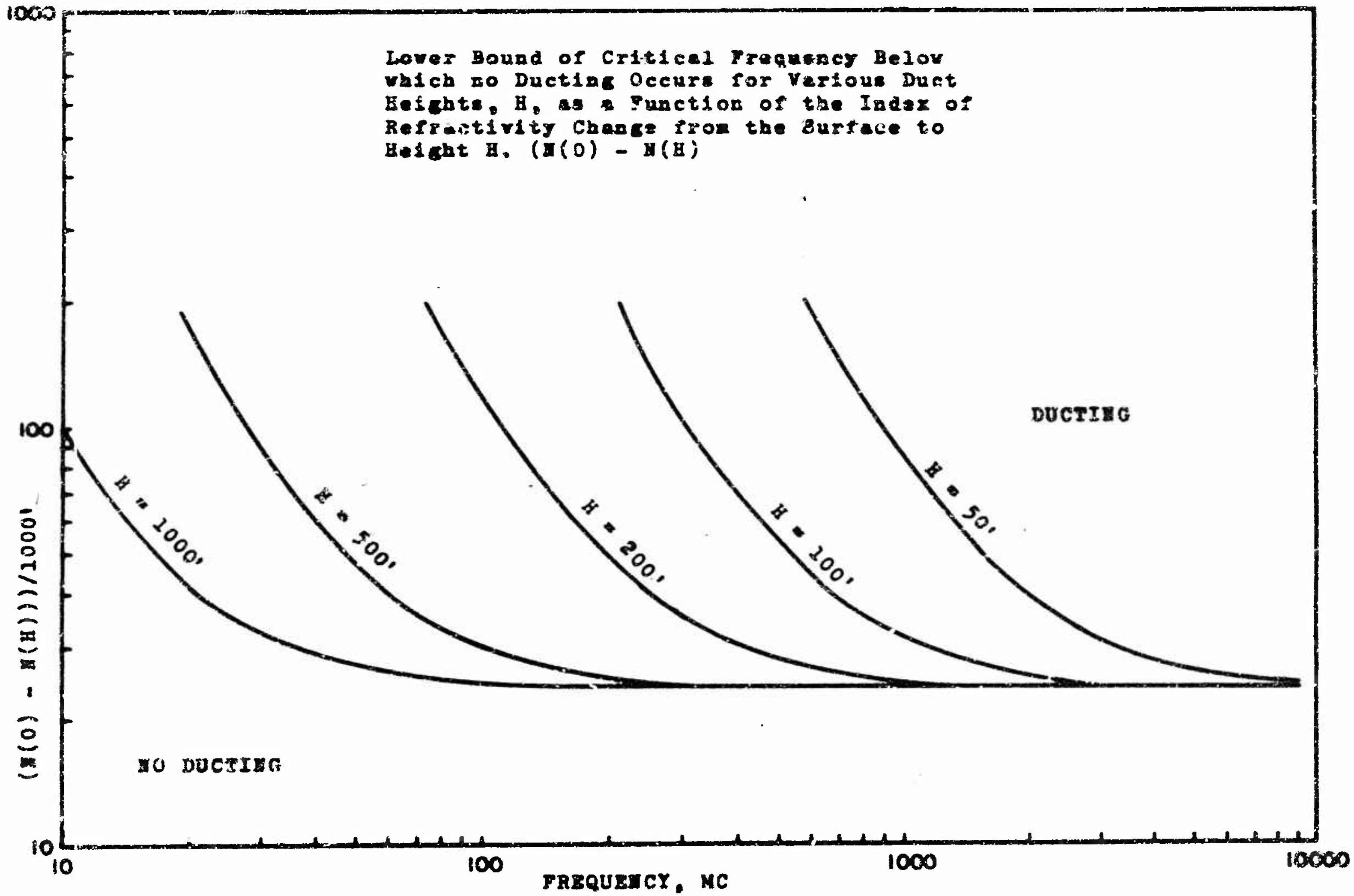


CHART 5. ATMOSPHERIC DUCTING

Source: Figure 1, ECOM-6009

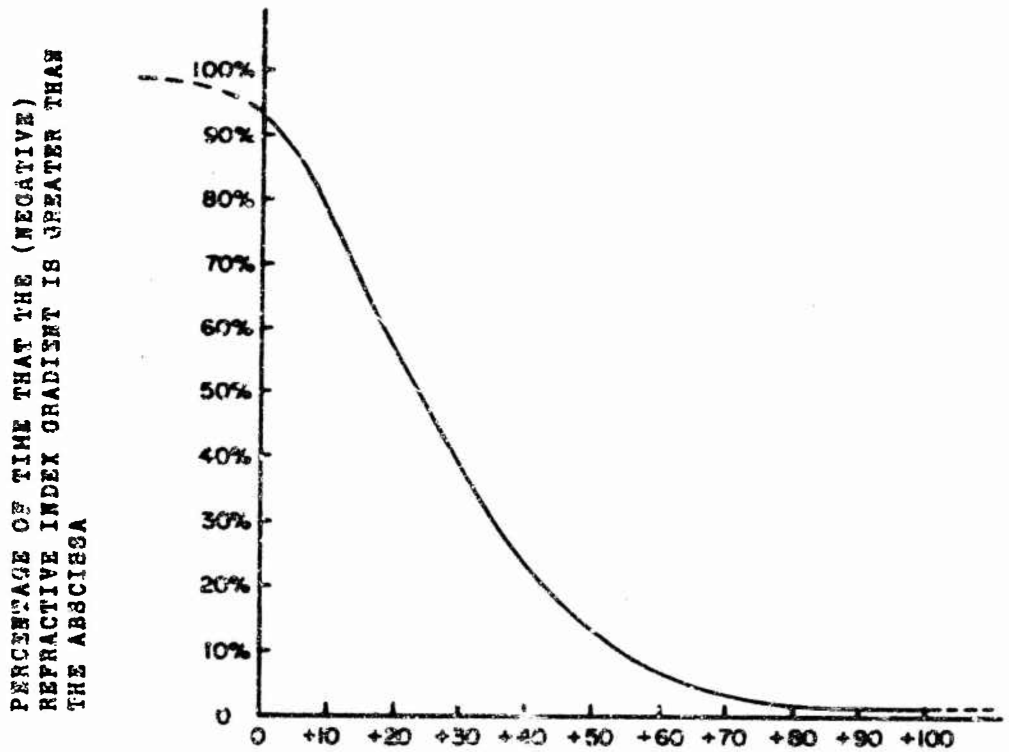


CHART 6. ATMOSPHERIC REFRACTIVITY CLIMATOLOGY FOR SOUTHERN ARIZONA

Source: Table III.1, USAERDAA-MET-7-64

LIFE-CYCLE MATERIEL FACTORS

AND THEIR APPLICATION TO THE PLANNING

AND OPERATING MANAGEMENT OF A FLEET OF VEHICLES

by

E. Rattner

Research Analysis Corporation
McLean, Virginia

ABSTRACT: The paper presents a computational procedure for determination of total fleet cost, from phase-in through "equilibrium" as a function of an age degrading vehicle performance parameter termed "worth." The latter is quantified as the product of a "success index" and an "obsolescence factor." The "success index" is the empirically determined probability of a tank, for example, being available and sufficiently reliable to complete a 50-mile movement at any time. The "obsolescence factor" represents the tank's mission performance capability on a scale of 0-1.0. Both factors, and hence the "worth index" are functions of the tank's age, measured in both calendar years and mileage.

This paper proposes an approach to the evaluation and costing of a fleet of combat vehicles. It considers a situation in which:

. A new-model combat vehicle fleet (e.g. the M60 tank) is being phased in to the combat inventory and displacing the earlier model fleet (Chart 1).

. Fleet size requirements have been determined for the new fleet and for each of its functional segments, the "subfleets" (Chart 2).

. The usage requirement for each segment has also been determined.

. A measure of vehicle performance effectiveness (a "Worth Index") is in use; it is primarily related inversely to usage (accrued mileage) and to a lesser degree related inversely to time (vehicle age) (Chart 3).

. The cumulative costs through the mileage life of a typical vehicle have been estimated.

1. What kind of relative effectiveness can we expect in the projected fleet?

2. What will the fleet operating cost be?

3. What will the cost be of establishing the projected fleet?

4. What rate of production will be necessary for replacements?

CHART 1

Fleet Life Cycle

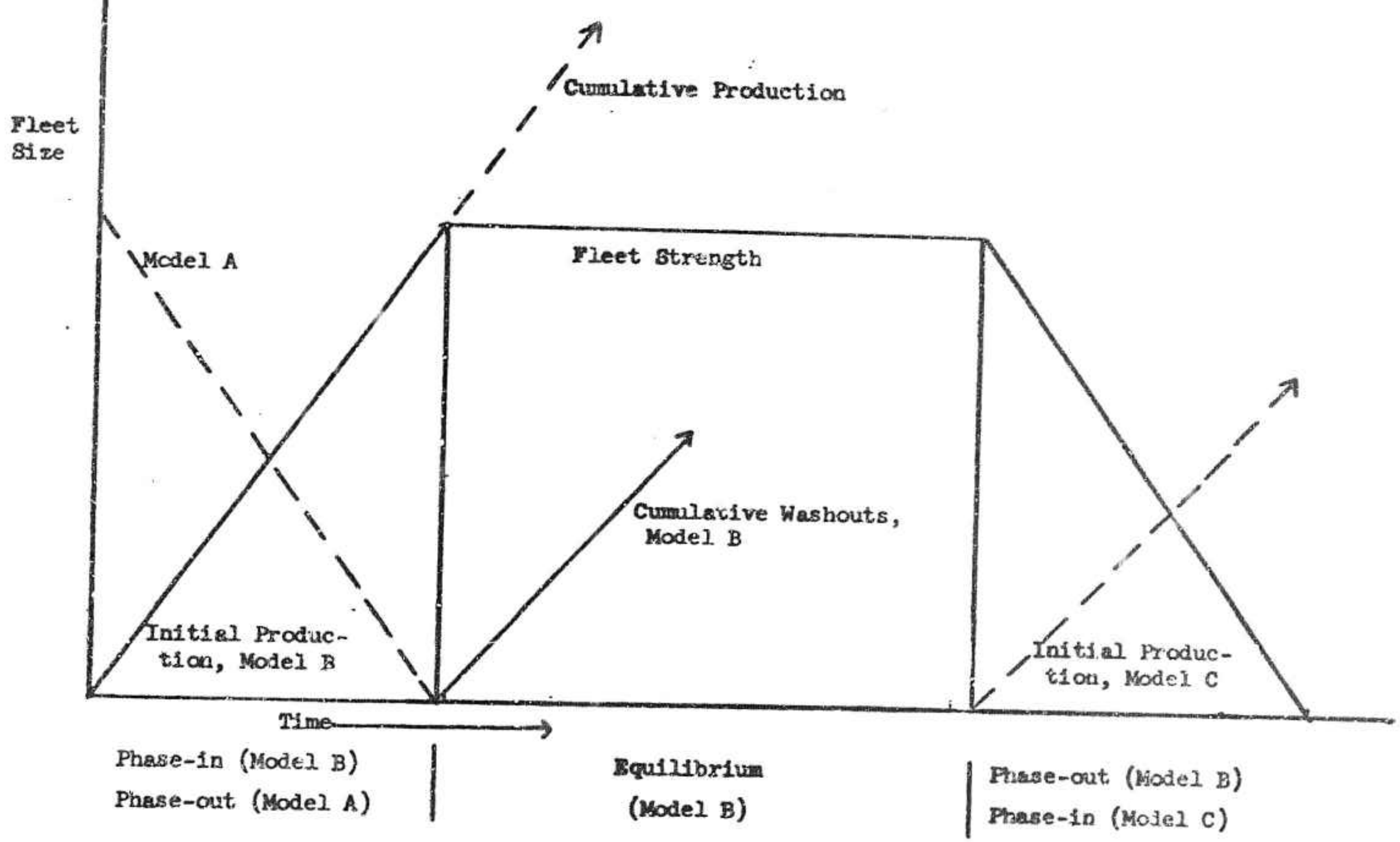


CHART 2

Balanced Fleet, Vehicle Flow

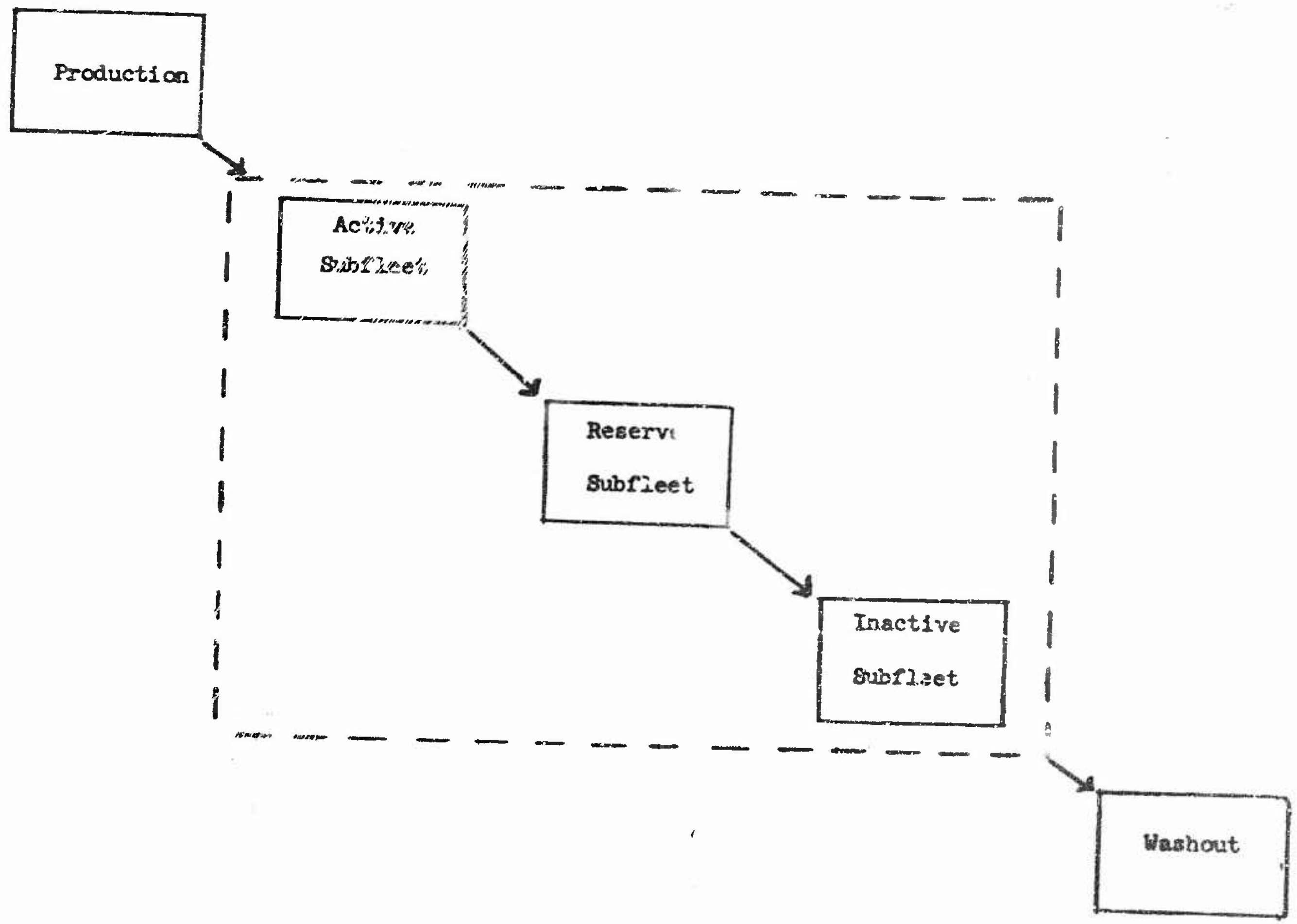
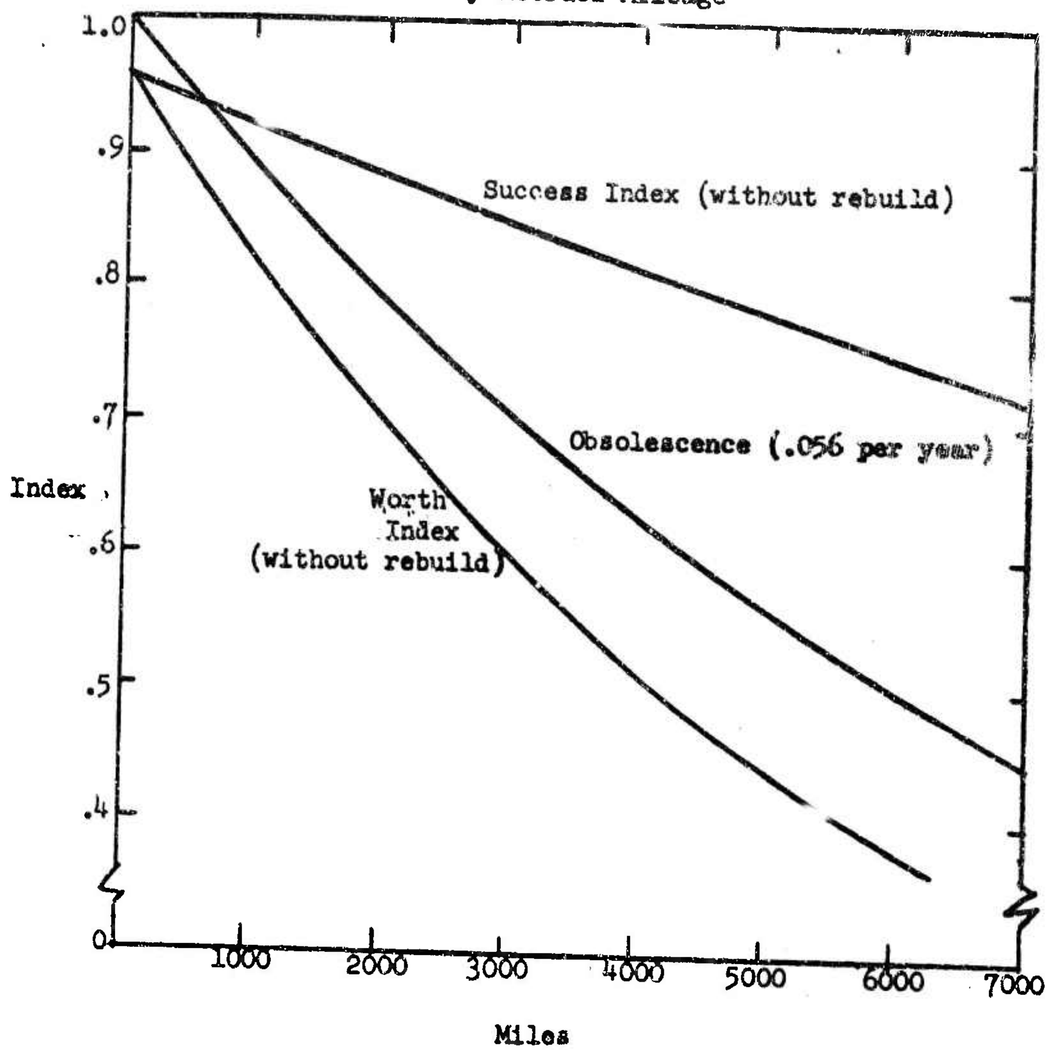


Chart 3

Indices of M60 Tank Performance,
By Accrued Mileage



THE FLEET PLANNING MODEL

General Approach

The fleet planning model is based on the life-cycle performance and cost pattern of the typical vehicle. By extension to a continuous-flow steady-state model, in which there are a great number of vehicles passing through the system, the characteristics and alternatives of the projected fleet will be considered.

The steady-state of the vehicle system begins at the conclusion of the phase-in period. The phase-in period represents the "investment" process in establishing an equilibrium "weapon system." Phase-in includes all the effort and cost involved in acquiring and operating a vehicle fleet to that point at which it can begin to operate on a steady-state basis. As a corollary, all costs incurred during the equilibrium phase are system operating costs.

The three major cost elements during the equilibrium period, acquisition cost, maintenance cost and rebuild cost, constitute categories of system operating cost (in contrast with their designations as factors of single vehicle costs). Vehicle acquisition during the phase-in period is an investment cost, an investment in achieving an equilibrium fleet. By the same token, the maintenance cost during the phase-in period would also be an investment cost, although in conventional economic usage, it would be considered as an operating cost.

The elements of this planning model, given a subfleet structure requirement and the usage annually for each subfleet, are (Chart 4):

1. Calculate the usage rate per vehicle for each subfleet.

2. Calculate the lower limit of the worth of the Active Subfleet (as the criterion for the entire fleet profile) as a function of age and mileage.

3. Select a minimum worth threshold for the Active Subfleet.

4. Calculate the vehicle flow rate that will provide this worth threshold.

5. Calculate the profiles of fleet worth and cost that the vehicle flow rate will provide.

These several steps can be shortened if the planner is given at the outset three sets of data:

1. The subfleet structure (V_{E_1}).*
2. The subfleet annual usage (μ_{1_t}).
3. The Active Subfleet minimum worth threshold (W_{E_1}).

* A List of Symbols is at the end of the paper.

Chart 4

Fleet Planning Factors and Requirements

1. Fleet Structure Requirements (V_{E_i})
   Training-Use Requirement (μ_{i_t})

2. Usage Rate Per Vehicle by Subfleet (U_i)

3. (Implied) Minimum Performance Standard (W_{E_i})

4. Vehicle Flow (Delivery) Rate (V_t)

5. Fleet Worth Profile $(W = a\, b^V)$

6. Fleet Cost Profile $(C_{T_{cum}} = C_A + C_{m_{cum}}\ [+ C_R])$

The flow rate can be calculated from these inputs, the values of a, b, and O are known from empirical data:

$$V_t = \frac{\mu_{1_t} \ (\log b) + {}^{V}E_1 \ (\log O)}{\log W_{E_1} - \log a}$$

The Mechanics of Vehicle Flow During Equilibrium

Use Patterns. A tank fleet is being operated in equilibrium; to observe the fleet operating pattern, consider a typical vehicle in flow through that system (Charts 5, 6, and 7). In Chart 5, "Hypothetical Use-Life for an M60 Tank," the new tank enters the Active Subfleet and begins to accrue mileage at a prescribed rate of usage, U_1. When its replacement arrives to enter the Active Subfleet, it will be at the point shown as M_{E_1} (mileage at the termination of its use within the Active Subfleet), which time point is designated as t_{E_1}. It then enters the Reserve Subfleet at a different (presumably lesser) rate of usage, U_2, and when it has acquired a given accrued mileage, M_{E_2} at time t_{E_2}, it is transferred to the Inactive Subfleet. In our example, we show it as accruing no mileage in the Inactive Subfleet ($U_3 = 0$) until the time point, t_{E_3}, when it washes out of the Inactive Subfleet as well as the total M60 fleet. The usage of this typical vehicle has its consequences in patterns of performance capability and cost.

Performance Patterns. In Chart 6, "The Hypothetical Performance Potential for an M60 Tank," the typical vehicle's performance-life is depicted, starting with a performance capability, measured in Worth Index values, close to 1.0 at the start of its fleet life. In time, as it wears out through use and is affected by obsolescence, its potential for reliable performance diminishes until it completes its service in the Active Subfleet. At that point in its life (shown as M_{E_1}), it is transferred from the Active to the Reserve Subfleet.

The lessening of its capability continues until it completes its Reserve Subfleet life at mileage M_{E_2}. At this point, under the usage depicted in Chart 5 (no additional mileage), the performance curve simply continues to descend, affected solely by obsolescence, until the mileage point shown in Chart 6 as M_{E_3} (time point is t_{E_3}). At the conclusion of its Inactive Subfleet life, the vehicle has reached its lowest point of capability. The dotted line shows a potential modification in this performance pattern that could be accomplished by a rebuild at some point about half way through its life. The rebuild mileage-point designated here as M_R, is the point at which the rebuild would be instituted for each vehicle flowing through the system.

CHART 5

Hypothetical Use-Life for an M60 Tank

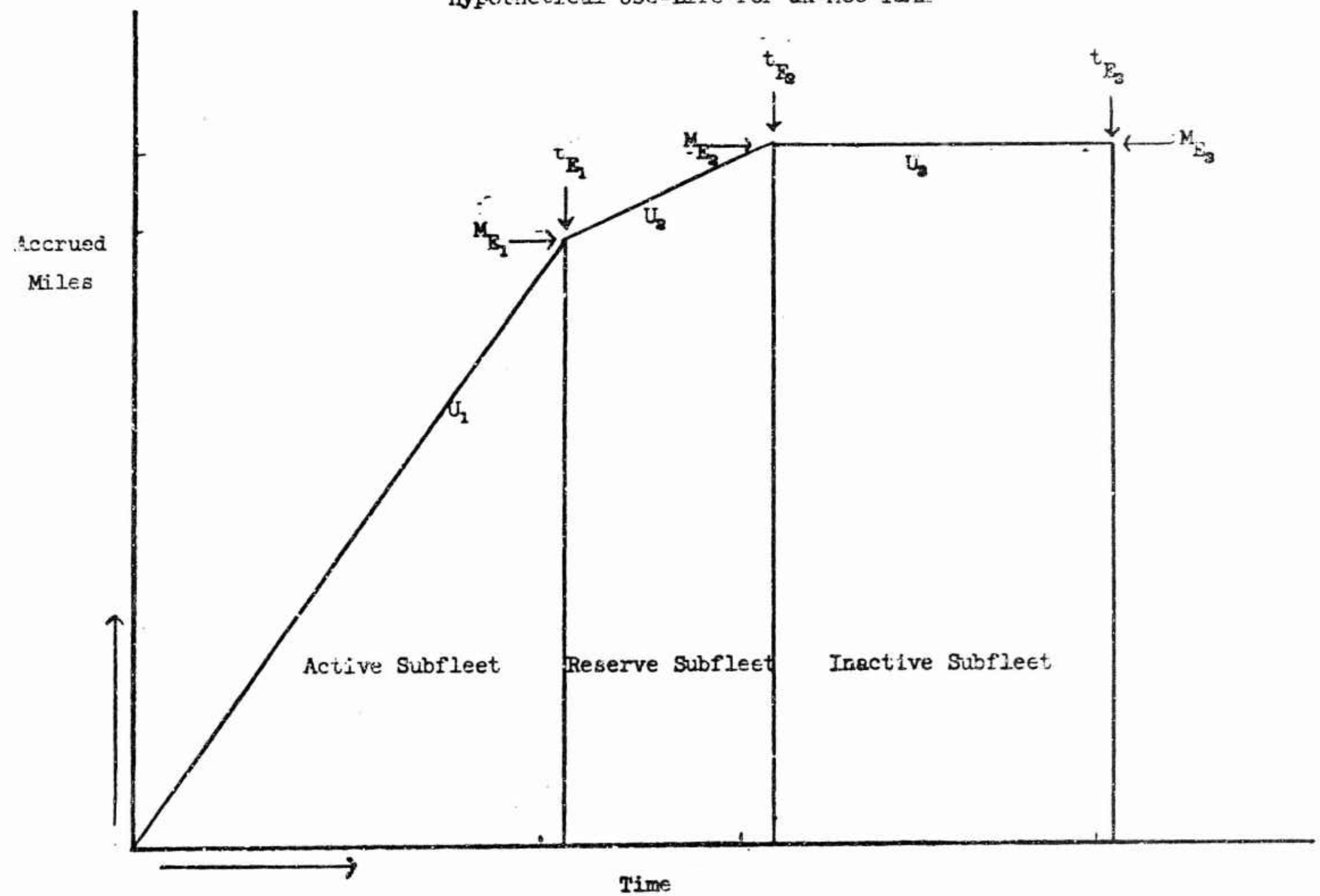
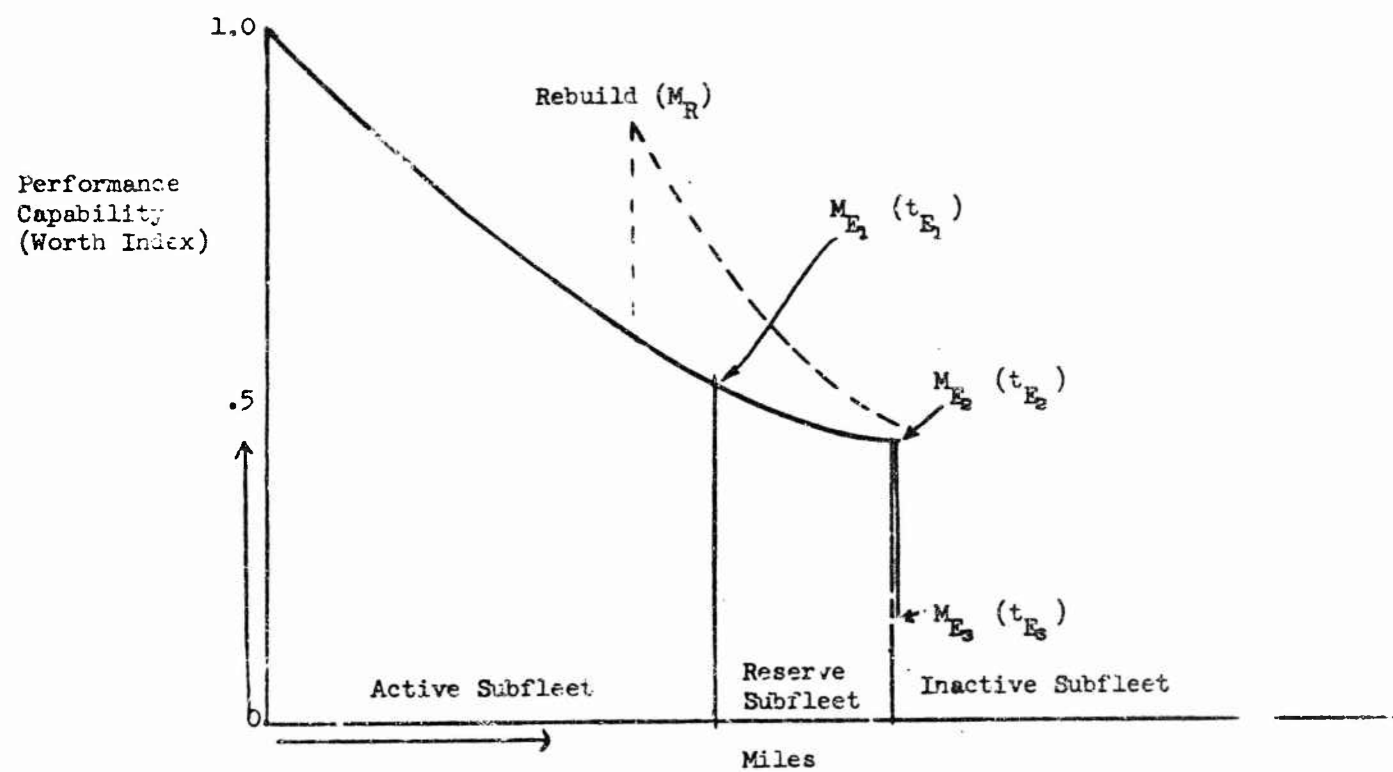


CHART 6

Hypothetical Performance Potential for an M60 Tank



The reliability of the vehicle can be largely restored by rebuilding. Under certain vehicle-use patterns, rebuild may be instituted more than once. In each instance of rebuild, there is both a gain in performance and an offsetting cost for that rebuild. Apart from pointing out in Charts 5 and 6 that a prescribed rebuild policy can modify both the performance and cost patterns of the typical vehicle, we will defer consideration of the effect of rebuild in order to avoid complicating the model at this point.

Cost Patterns. The cost pattern of the same typical vehicle (Chart 7), shows a very large acquisition cost, maintenance costs accruing through extended use at an increasing cost rate until the completion of the vehicle's Active Subfleet life, M_{E_1}, with a continuing increase in the maintenance cost rate through its Reserve Subfleet use. The rebuild, whose performance effect was shown in Chart 6, seen here as a rebuild cost (C_R) instituted during the latter half of the Active Subfleet life, is a large immediate cost increment somewhat offset by the lower cost for maintenance for a period subsequent to rebuild. We have made an assumption that the maintenance cost of a vehicle in inactive status for some several years is negligible (i.e., zero cost) compared to its previous operating cost.

The three charts depict the hypothetical experience of a single vehicle flowing through a combat vehicle system in equilibrium. Depending upon the two key parameters of fleet structure (i.e., fleet and subfleet size) and the training-use requirement, these curves vary. The usage rate per vehicle determines the costs directly and the performance inversely.

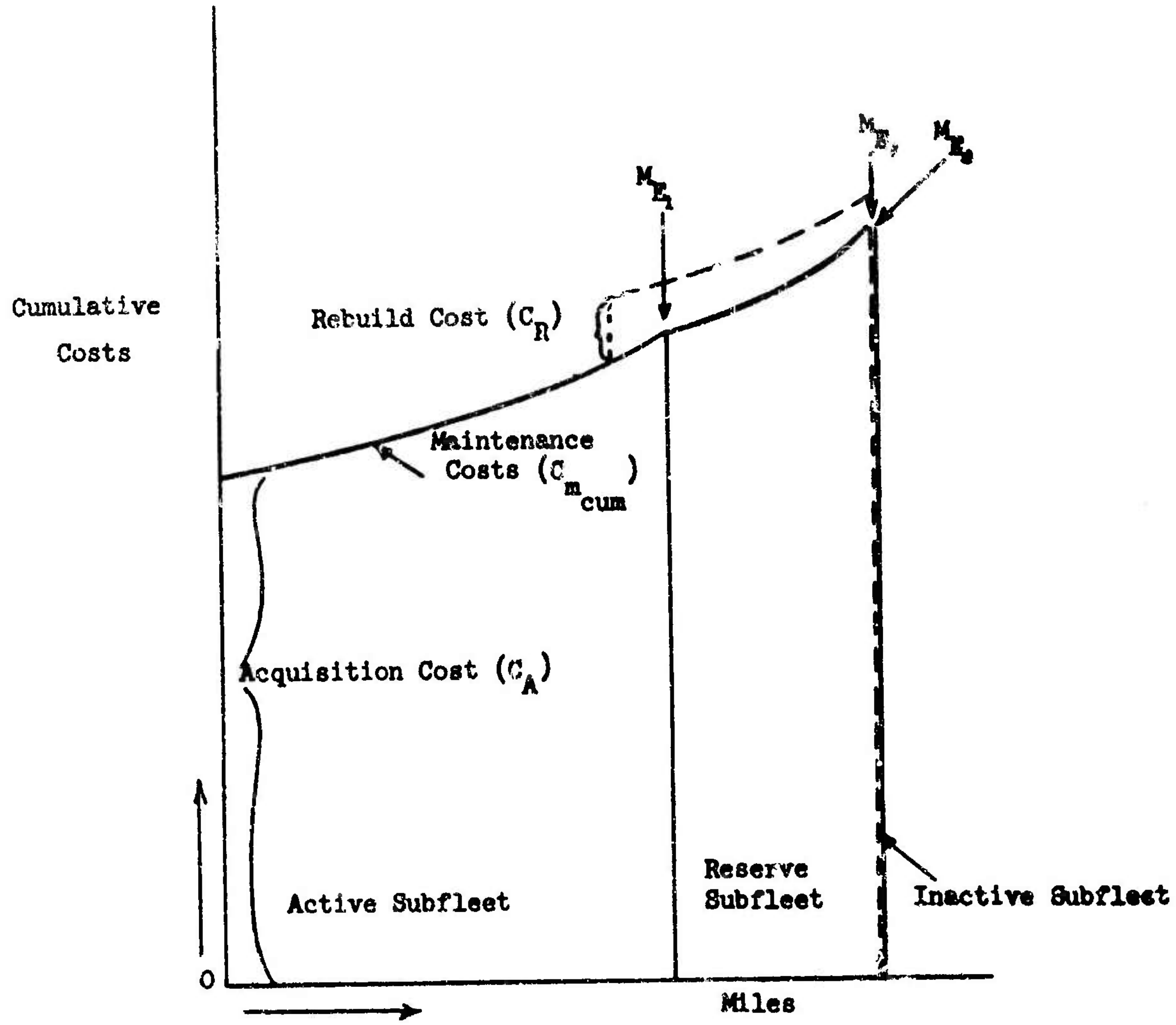
Fleet Operations As a Flow Process

Background. The operation of the balanced fleet can be viewed simply as the flow of evenly-spaced vehicles through these use, performance, and cost patterns. For example, if we assume that the gain-and-loss rate of this equilibrium system were 365 vehicles per year, any given vehicle at some point in the system would follow its immediate predecessor in the system by one day and lead its successor by one day. Each of these three randomly-located vehicles would have assessable performance capabilities depending upon its mileage accrual at that point in time. Similarly, each would have generated cumulative costs described by the cost pattern curves.

No two vehicles within the same organizational unit run exactly the same number of miles, even on a common activity as a training exercise. While it is true that mileage is not accrued as continuously and evenly as we assume here, yet for large groups of vehicles, and we are attempting here to predict for a fleet and not for a single vehicle, this depiction of the expected experience of a single vehicle is the average of all such vehicles in the fleet.

CHART 7

Hypothetical Cost Pattern for an M60 Tank



Major Factors in Fleet Planning.

Operational Patterns. The primary determinant of vehicle performance and of costs is usage, as measured by accrued mileage. For the purpose of further analysis, we can convert the performance potential and cost pattern charts into functions of time or of numbers of vehicles. The advantage would be to show the influence of time on certain performance and cost levels, and similarly, the influence of the flow rate of vehicles through the system can be analyzed more readily by converting mileage into vehicular values.

The two basic factors which are needed to initiate our fleet plan are fleet structure (subfleet sizes) and the use requirement. The latter factor is the aggregate number of miles of use for each subfleet annually.

Consider a hypothetical tank fleet, Fleet X, with its Active, Reserve, and Inactive Subfleets, respectively X_1, X_2, and X_3. Assume that the fleet size requirement (V_E) is 5,000 vehicles with the respective subfleet sizes, (V_{E_1} =) 1800, (V_{E_2} =) 1200, and (V_{E_3} =) 2000, and further assume that the required operating rates (μ_{i_t}) for these subfleets is 900,000 miles annually for Subfleet X_1, 100,000 miles annually for Subfleet X_2, and 0 mileage for Subfleet X_3, a total of one million miles annual use. Symbolically, the first set of data are the V_{E_i} values and the second set are the U_i values. The annual operating rate per vehicle (U_i) for each subfleet is then,

$$U_i = \frac{\mu_{i_t}}{V_{E_i}}$$

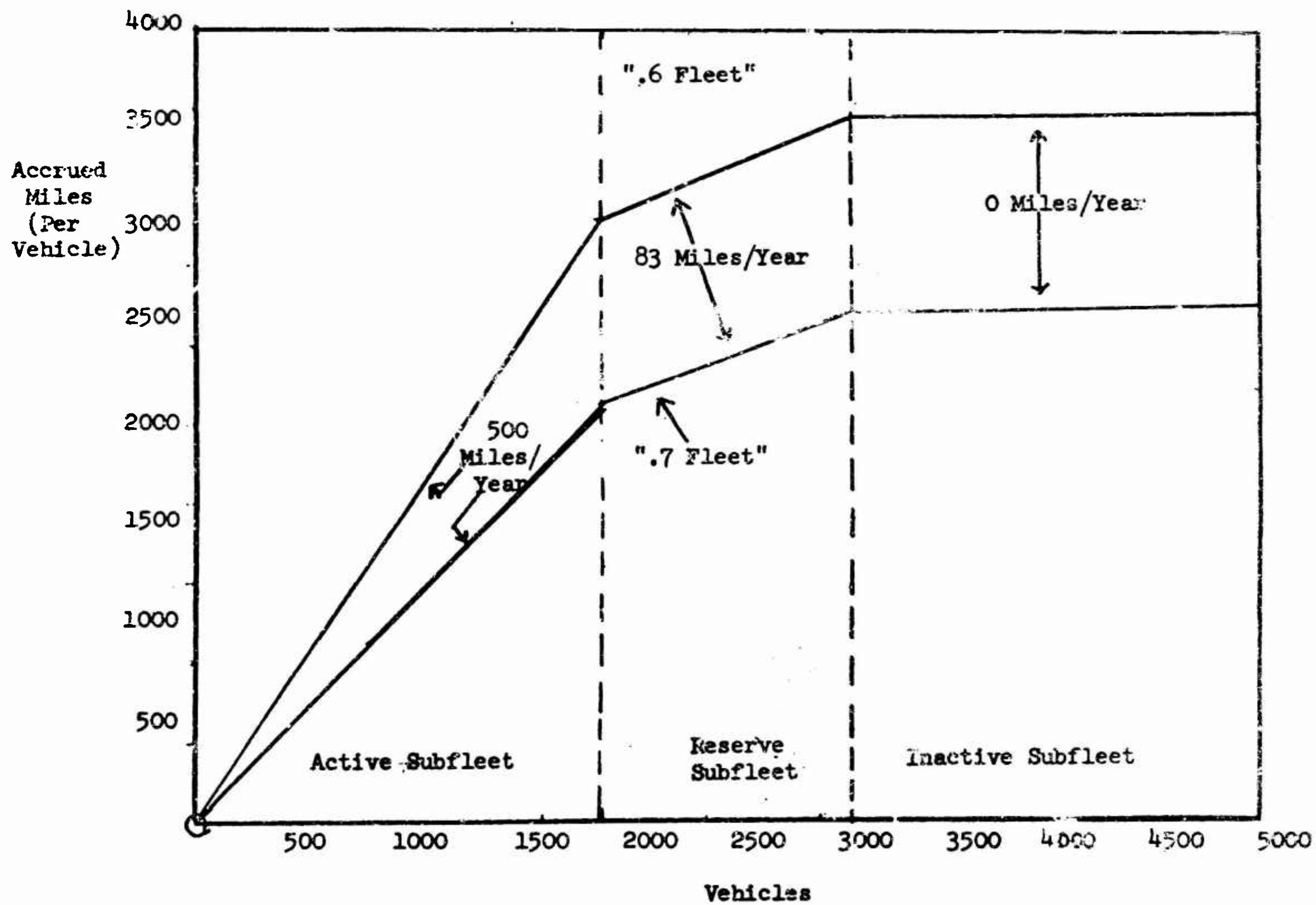
making the respective subfleet values, U_1 = 500 miles per year; U_2 = 83 miles per year; and U_3 = 0 miles per year.* U_i is determined by the requirements, and since U_i is equal to the mileage accrued by a vehicle in a given fleet, divided by its lifetime in that fleet,

$$U_i = \frac{M_{E_i}}{t_{E_i}}$$

* See Chart 8.

CHART 8

Usage Patterns of Two Alternative Vehicle Fleets



at a constant rate of use, the mileage accrued during the vehicle's period of life in a given subfleet is directly related to the time spent within that subfleet. If we determine the subfleet lifetime (t_{E_i}), we can prescribe the accrued mileage limit for that subfleet's vehicles (M_{E_i}), and vice versa.

One simple logistical policy based upon this fact is to prescribe an indefinite lifetime within the subfleet so as to "squeeze" as much mileage as possible from a given vehicle in its subfleet life. Obviously, these vehicles do not have an indefinitely long lifetime because of their depreciating worth through operating use and time.

Worth. Subfleet lifetime is a function of the minimum standard of performance that we will accept for the subfleet. It is a function of such a standard whether we recognize the existence of any standard of performance or not. The performance measure that we have referred to, the Worth Index, can be applied as the basis for determining the mileage life and time for our subfleets. Using a minimum performance capability of the Active Subfleet as our limiting criterion, lifetime and accrued mileage for Subfleet X_1 can be determined from the Worth Index.

$$W = ab^{M}0^{t}$$

$$W = .964(.999959)^{M}\ (.944)^{t}$$

If we establish as a minimum performance standard the worth of the least effective vehicle in the Active Subfleet, we can compute the age and mileage limits that each standard would impose. For example, for a minimum worth level of .5, a vehicle would be 8.4 years in the Active Subfleet. This is the worth of the oldest vehicle that has passed through the Active Subfleet and is to be transferred into the Reserve Subfleet. If we accept a minimum Active Subfleet worth of .6, lifetime in the Active Subfleet would be 6.1 years; for a worth of .7, its Active Subfleet lifetime would be 4.1 years (see Table 1).

The value of the minimum worth determines the lifetime, and conversely, the lifetime determines the minimum worth level within the subfleet. Determination of the acceptable minimum performance standard is not the function of the logistical planner, but he can feedback the cost and effectiveness implications of alternative standards to the requirements staff.

If the fleet had a .6 worth minimum for its Active Subfleet, then t_{E_1} would be 6.1 years and at that time a vehicle will have accrued a mileage of 3,037 miles. What is the rate of replacement (V_t) for this ".6 Fleet"? The rate of flow of vehicles, as we have stated, is constant for all parts of the system.

TABLE 1

VEHICLE WORTH AS A FUNCTION OF MILEAGE AND TIME

(Age and Mileage at Transfer from Subfleet X_1 to Subfleet X_2 for Selected Worth Values)

W_{E_1}	t_{E_1}	M_{E_1}
.5	8.4100 (years)	4205 (miles)
.6	6.0738	3037
.7	4.1003	2050
.8	2.3894	1195

Based upon: $W = ab^M \, O^t$

$$= .964\,(.999959)^M\,(.944)^t$$

$$V_t = \frac{V_{E_1}}{t_{E_1}} = \frac{1800}{6.1} = 296 \text{ vehicles per year}$$

or

$$V_t = \frac{U_1}{M_{E_1}} = \frac{900,000}{3,037} = 296 \text{ vehicles per year}$$

The flow rate of the ".6 Fleet" is 296 vehicles per year. Through similar computations, we can complete the equivalent parameters for subfleets X_2 and X_3. The typical vehicle would have accrued 3,037 miles when it left the active fleet (M_{E_1}). It would enter the Reserve Subfleet for an additional four years, accruing an additional 337 miles. It would then be assigned to the Inactive Subfleet (without further mileage) for 6.7 additional years. This would be a total fleet life of 16.9 years with a total accrued mileage of 3,374 miles. The average annual usage for the total fleet experience would be 200 miles per year per vehicle.

What are the implications for the performance capability of the three subfleets under this operating use and replacement rate? As we have said, one of the bases for this fleet was a minimum Active Subfleet standard worth of .6; the performance capability of Subfleet X_1 would range from the high capabilities of its newest vehicles down to a .6 worth for those vehicles about to transfer into the Reserve Subfleet. The best vehicle in Subfleet X_2 would therefore have .6 worth, and the oldest vehicles in Subfleet X_2 would have a worth of .469. The Inactive Subfleet worth range would be from .469 down to .318 for those vehicles ready for final washout.

Suppose we were to select a minimum standard of .7 for the Active Subfleet instead of .6. The Fleet Plan Table, Table 2 shows the contrasting characteristics of the alternative total fleets.

The Fleet Plan as a Function of the Vehicle Flow Rate. Up to this point, we have made the system primarily a function of mileage and time. These key factors are convertible to vehicle quantities to depict fleet profiles of effectiveness and cost. This transformation can readily be made from the basic equations in which mileage and time are the key variables. For example,

$$U_1 = \frac{M_{E_1}}{t_{E_1}} = \left(\frac{M_1}{t_1}\right) \text{ and } t_{E_1} = \frac{V_{E_1}}{V_t} \left(\text{or } t_1 = \frac{V_1}{V_t}\right)$$

$$M_1 = \frac{U_1}{V_t} \quad V_1$$

Table 2

FLEET PLAN TABLE
Parameter Values (Fleet X)
<u>Hypothetical Data</u>

	".6 Fleet"				".7 Fleet"			
Item	X	X_1	X_2	X_3	X	X_1	X_2	X_3
1 μ_{i_t}	1,000,000	900,000	100,000	-	1,000,000	900,000	100,000	-
2 V_{E_i}	5,000	1,800	1,200	2,000	5,000	1,800	1,200	2,000
2a (cum)		(1,800)	(3,000)	(5,000)		(1,800)	(3,000)	(5,000)
3 U_i	200	500	83.3333	-	200	500	83.3333	-
4 M_{E_i}	3374.3333	3036.9000	337.4333	-	2277.9417	2050.1500	227.7917	-
4a (cum)			(3374.3333)	(3374.3333)			(2277.9417)	(2277.9417)
5 t_{E_i}	16.8717	6.0738	4.0492	6.7487	11.3897	4.1003	2.7335	4.5559
5a (cum)			(10.1230)	(16.8717)			(6.8338)	(11.3897)
6 W_{E_i}	.318	.6	.469	.318	.456	.7	.593	.456
7 V_t	296.3548	296.3548	296.3548	296.3548	438.9923	438.9923	438.9923	438.9923
8 C_{A_t}	$29,635,480				$43,899,230			
9 C_{m_t}	$ 7,006,198	$6,098,389	$909,809		$ 6,180,573	$5,378,534	$802,039	
10 C_{R_t}	$ 6,519,806				$ 9,657,831			
11 $C_{m_{cum}}$	$ 23,648	$ 20,578	$ 23,648	$23,648	$ 14,079	$ 12,252	$ 14,079	$14,079
11a (increment)			3,070	-			1,827	
12 C_{m_M}	$7.01	$6.78	$9.10	-	$6.18	$5.98	$8.02	-

in the Active Subfleet, the usage rate (U_i) equals the per vehicle accrued mileage (M_{E_i}) in Subfleet X_i divided by the time (t_{E_i}) spent in the subfleet; also, t_{E_i} equals the Subfleet X_i vehicle quantity (V_{E_i}) divided by the replacement rate (V_t). Any point i in the mileage, time, and vehicle ranges of the Active Subfleet is identically related.

Worth. The Worth Index, based on mileage and time,

$$W_{E_i} = ab^M (.944)^t = ab^{M_{E_i}} (.944)^{t_{E_i}}$$

becomes, in vehicle values,

$$W = ab^{\left\{\frac{U_i V_i}{(V_t)}\right\}} (.944)^{\frac{(V_i)}{(V_t)}} \quad \left| \begin{array}{l} V_i = V_{E_i} = 1800 \text{ (exit from Fleet } X_i\text{)} \\ V_i = 0 \text{ (entry to Fleet } X_i\text{)} \end{array} \right.$$

and since U_i and V_t are constants, Worth is now a function of V_i, any vehicle i in the Active Subfleet.

The same equation can be used to show the effect on a given vehicle's worth of the flow rate by making U_i and V_i the constants and allowing V_t to vary. From the form of the equation, for any vehicle in the system, the greater the flow rate, the greater the worth.*

Costs. The cost of equilibrium fleet operation is

$$C_T = C_A + C_m + C_R$$

The total cost is equal to the acquisition cost plus the maintenance cost plus the cost for rebuild, if a rebuild policy is applied. On an annual basis, symbolically, this would be

$$C_{T_t} = C_{A_t} + C_{m_t} + C_{R_t}$$

The total annual cost for the fleet is equal to the annual acquisition cost plus the annual maintenance cost plus the annual rebuild cost. The data that has been collected in previous studies has indicated a close fit for the cumulative maintenance costs per vehicle of

$$C_{m_{cum}} = aM^b = .52154M^{1.319802}$$

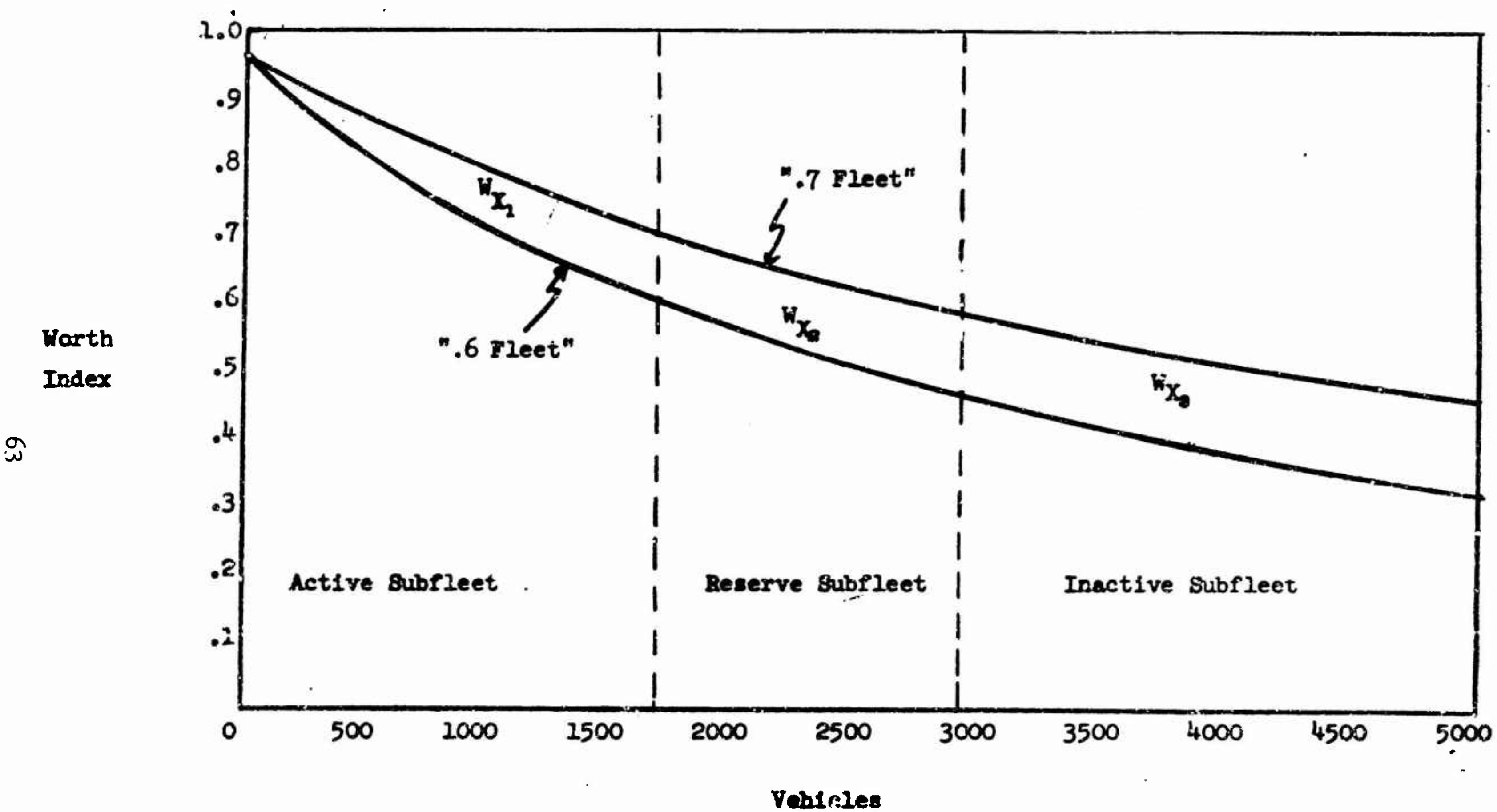
This cumulative function for maintenance costs, the major variable in operating costs, can be estimated as a function of the accrued mileage of a tank. In the example we are using for Subfleet X_1, with a

---

* See Chart 9.

CHART 9

Worth Profiles of Two Alternative Vehicle Fleets



minimum worth cutoff of .6, the vehicle which will have reached that point of mileage (and worth) will have accrued \$20,578 of maintenance costs. It will have accrued an additional \$3,070 in its Reserve Subfleet use, making the total maintenance cost cumulation when it enters the Inactive Fleet, \$23,[illegible]8.

The annual maintenance for each of the subfleets in the ".6 Fleet," shown in the Fleet Plan Table, would be \$6.1 million for the Active Subfleet and \$.9 million for the Reserve Subfleet with assumed negligible cost for maintenance during Inactive Subfleet life. Annual maintenance cost for the .6 Fleet would be \$7.0 million, the annual acquisition cost (at a hypothetical \$100,000 per vehicle) would be \$29.6 million, making the total cost, under a no-rebuild policy, \$36.6 million to sustain a fleet of this structure at the .6 minimum worth level.* A policy of one-rebuild per vehicle would add \$6.5 million annually to the total (at an assumed rebuild cost of \$22,000 per vehicle).

The equivalent estimates for the higher performance ".7 Fleet" are \$6.2 million for annual maintenance, slightly less than for the .6 Fleet, and annual acquisition costs of \$43.9 million, making its total under a no-rebuild policy \$50.1 million. A one-rebuild per vehicle policy would add \$9.7 million annually.

Fleet Phase-in Costs

The cost of creating the equilibrium fleet is its phase-in cost. This is the period during which the first vehicle delivered into Fleet X has moved from the initial vehicle position through each of the intervening 4998 vehicle-position to the final vehicle position, V_{E_3}.

At the moment when it washes out of the fleet, the equilibrium phase begins. The cost of phase-in is then the aggregate cumulative cost of each of the 5000 vehicles at that moment.

$$\text{Phase-in Cost} = V_{E_3} (C_A + C_R) - V_R C_R + C_m$$

$$C_m = \sum_{i=0}^{5000} C^{**}_{m_{cum_{V_i}}} = \int_0^{1800} 1.04017(v)^{1.319802} + \int^{3000} 2673.82(V)^{.27256} + \int_{3000}^{5000} 23,648$$

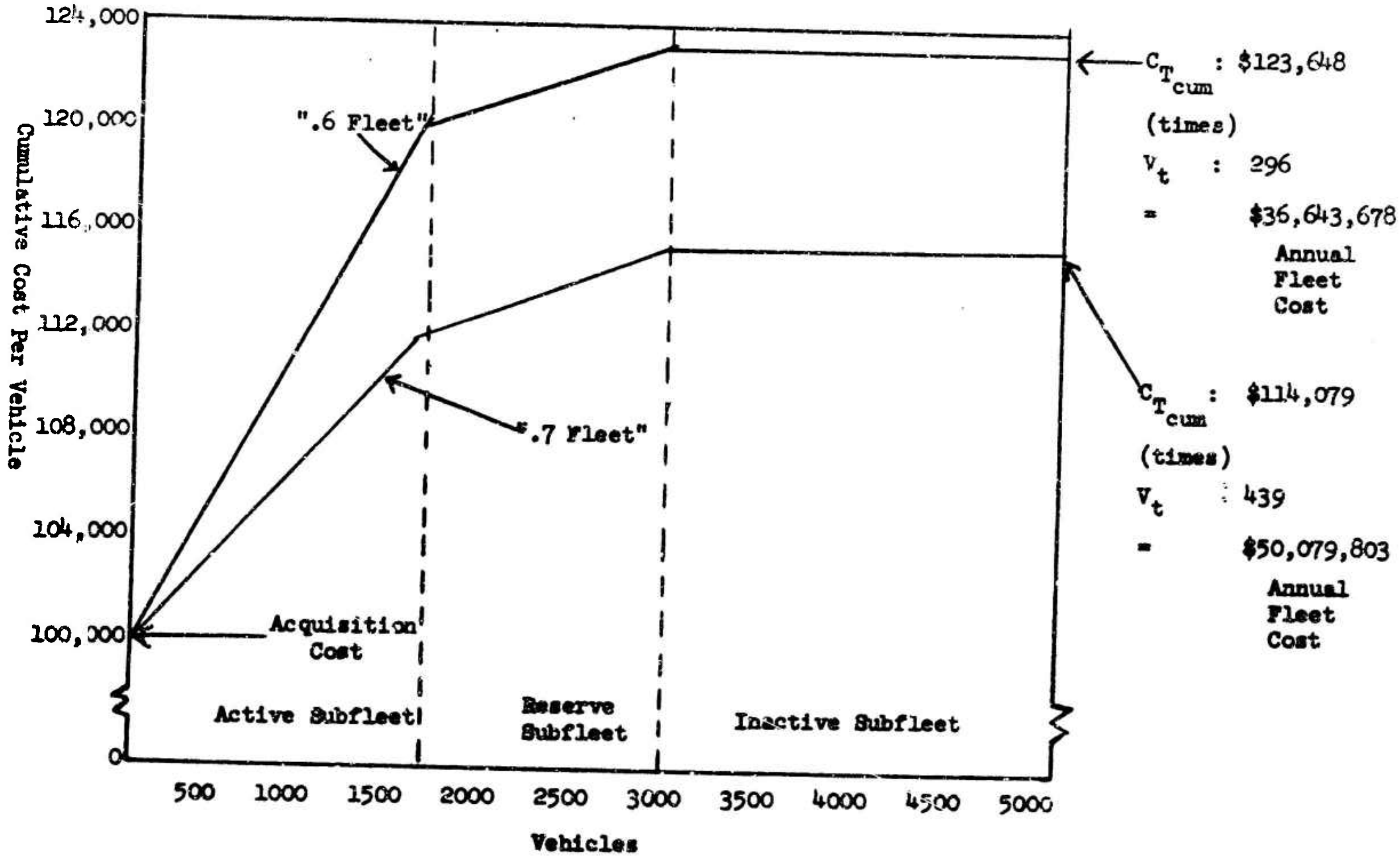
$$= \frac{1.04017}{2.319802}\left[V^{2.319802}\right]_0^{1800} + \frac{2673.82}{1.272256}\left[V^{1.272256}\right]_{1800}^{3000} + 23,648\,[V]_{3000}^{5000}$$

* See Chart 10.

** Maintenance costs for a ".6 Fleet" during phase-in; this is a vehicle-dependent transform of the equation for $C_{m_{cum}}$ described previously.

CHART 10

Cost Profiles of Two Alternative Vehicle Fleets



$$= .44839 \left[V^{2.319802} \right]_{0}^{1800} + 2101.6368 \left[V^{1.272256} \right]_{1800}^{3000}$$

$$+ 23,648 \, [V]_{3000}^{5000}$$

$$= (.44839)\,(1800)^{2.319802} + 2101.6368 \left[(3000)^{1.272256} - (1800)^{1.272256} \right]$$

$$+ (23,648)\,(2000)$$

$$= (.44839) \text{ antil } [2.319802 \,(\log 1800)\,] + 2101.6368 \,[\text{antil } \{1.272256$$

$$(\log 3000)\} - \{\text{antil } 1.272256 \,(\log 1800)\}\,]$$

$$+ 47,296,000$$

$$= (.44839)\,(\text{antil } 7.55158) + 2101.6368 \,[\text{antil } 4.42379 - \text{antil } 4.14154\,]$$

$$+ 47,296,000$$

$$CM = 15,967,463 + 26,649,479 + 47,296,000 = \$89,912,942$$

Phase-in Cost $= V_{E_3} \, (C_A + C_R) - V_R C_R + C_m$

$= 5000 \, (100,000 + 22,000) - (1500)(22,000) + 89,912,942$

$= 589,912,942$ (for a ".6 Fleet," no-rebuild policy)

$= \$666,912,942$ (for a ".6 Fleet," under a one-rebuild policy applied to each vehicle after it has reached the 1500th vehicle position)

Recapitulation

At the beginning of this paper, I posed several questions which a planner would be faced with in making logistical and operating decisions for a fleet of combat vehicles. I have tried to demonstrate through the perspective of an equilibrium fleet how such answers affecting logistical decisions might be arrived at.

1. What kind of relative effectiveness can we expect in the projected fleet?

Relative Effectiveness

	Fleet	Total	Active	Reserve	Inactive
Size		5000	1800	1200	2000
Worth Aggregate	(.6)	2797	1382	638	777
" "	(.7)	3301	1485	774	1042
Worth Range	(.6)	1.0-.32	1.0-.6	.6-.47	.47-.32
" "	(.7)	1.0-.46	1.0-.7	.7-.59	.59-.46
Mean Worth	(.6)	.56	.77	.53	.39
" "	(.7)	.66	.83	.65	.52

2. What will the fleet operating cost be?

Annual Fleet Operating Cost
Millions of Dollars

Fleet	Total	Active	Reserve	Inactive
(.6)	36.6	35.7*	.9	--
(.7)	50.1	49.3*	.8	--

* Acquisition Cost allocated to Active Subfleet entirely.

3. What will the cost be of establishing the projected fleet?

Phase-in Cost*

Fleet	(.6)	$589.9 Million
"	(.7)	$553.5 Million

* Under a no-rebuild policy.

4. What rate of production will be necessary for replacements?

Annual Replacement Rate

Fleet	(.6)	296 Vehicles
"	(.7)	439 Vehicles

I have applied the concept of a fleet of vehicles being operated on a policy of predetermined constant-level usage and replacement. These determinations have been based upon stated requirements of subfleet sizes and usage rates to which an empirical measure of effectiveness, the Worth Index, and expected cost for each vehicle passing through the system have been applied.

Application of this method to Army vehicle fleets is obviously dependent upon the accuracy and reliability of the data. The Worth Index that was used in this paper was based on a field test of the mobility of the tank; it is recommended that similar data be collected among a stratified sampling of vehicles to determine the proportion of each age stratum that will respond successfully with all systems in ready condition at the mission location. The cumulative operating cost for the typical vehicle through its lifetime are the second category of data essential to the method. Such data are now increasingly available on major equipment items reported by the Army's equipment reporting system.

Some expedient shortcuts were made with these data and the analysis in order to focus on the essentials of the concept. In the practical application, each of these deliberate omissions should be considered and included. Among these are:

1. The per vehicle costs of delivery from the factory to the acquisition unit, POL, and salvage value.

2. The phase-in and operating costs of maintenance and overhaul floats as specialized "subfleets."

3. The actual cost and condition of a vehicle deactivated through an extensive period of time.

4. The worth and cost profile of an overhauled vehicle.

Summary

Combat fleet requirements are currently in terms of numbers of vehicles without regard to qualitative differences. No distinction is made among the vehicles of the fleet, although it is recognized that, in general, the older vehicles of a fleet are less reliable. Cost and performance estimates can be made of every vehicle in the operating fleet by the performance model (the Worth Index) and the cost model presented here. Vehicle use and lifetime data should be analyzed to determine their effects on the qualitative and economic characteristics of the operating fleet.

Utilizing data of costs, fleet size requirements, and indices of reliability and worth, the paper's concept of a continuous flow of replacement vehicles describes the determination of delivery rates, operating rates, subfleet phasing, and washout. The paper considers the M60 main battle tank; it is also applicable to planning other tracked-and wheeled-vehicle fleets. It is intended to show how planning and management staffs may project dates of fleet buildup, cost cumulations through the end of fleet phase-in, and the cost and reliability and worth profiles of the fleet in equilibrium (constant level of input and output of the fleet).

Conclusions

1. It is feasible to create a combat vehicle fleet meeting specified criteria of reliability and performance for an indefinite period of years.

2. With the application of a specified minimum standard of performance (e.g., the Worth Index), fleet capability is directly determinable by the replacement rate.

3. Cost effectiveness of a combat vehicle fleet must be constrained by a minimum performance threshold for the fleet; cost effectiveness level of a combat vehicle fleet is met by minimization of the costs for a fleet of established standard.

4. Data should be developed to provide the measures of worth and cumulative cost throughout the range of accrued mileage-and-time for a representative sampling of the fleet.

LIST OF SYMBOLS

X	Fleet Designation
X_i	Subfleet i of Fleet X
X_1	Active Subfleet
X_2	Reserve Subfleet
X_3	Inactive Subfleet
μ	Fleet Mileage
μ_i	Subfleet X_i's Mileage
μ_{i_t}	Subfleet X_i's Annual Mileage
M	Miles (Single-Vehicle)
M_i	The ith Mile
M_{E_i}	The (Odometer) Mileage of a Vehicle Exiting from Subfleet X_i
V	Vehicle
V_i	The ith Vehicle
V_{E_i}	The Final (Oldest) Vehicle in Subfleet X_i; The Number of Vehicles in Subfleet X_i
V_t	Vehicle Flow Rate; Replacement Rate; Washout Rate
t	Time (in years)
t_i	The ith Time-point
t_{E_i}	The Elapsed Time (Since Entry into Fleet X) of a Vehicle Exiting from Subfleet X_i
t_E	The Elapsed Time (Since Entry in Fleet X) of a Vehicle Exiting from Fleet X
U	Usage Rate (in miles)
U_i	Usage Rate in Subfleet X_i
S	Success Index (Value)
S_i	Success Index Value of the ith Vehicle
S_{E_i}	Success Index Value of a Vehicle Exiting from Subfleet X_i

O	Obsolescence
O_{E_i}	Obsolescence Factor of a Vehicle Exiting from Subfleet X_i
W	Worth Index (Value)
W_i	Worth Index Value of the ith Vehicle
W_{E_i}	Worth Index Value of a Vehicle Exiting from Subfleet X_i
a, b	Conventional Mathematical Constants
$\Big\vert_f^g$	Range Limits of a Variable ($f \leq$ Variable $\leq g$)
C	Cost
C_T	Total Cost
C_A	Acquisition Cost
C_m	Maintenance Cost
C_R	Rebuild Cost
C_{T_t}	Annual Total Cost
C_{A_t}	Annual Acquisition Cost
C_{m_t}	Annual Maintenance Cost
C_{R_t}	Annual Rebuild Cost
$C_{m_{cum}}$	Cumulative Maintenance Cost (for a Single Vehicle)
$C_{T_{cum}}$	Cumulative Total Cost (for a Single Vehicle)
C_{m_M}	Maintenance Cost Rate (Per Mile)
$\bar{C}_{m_M}$	Mean Maintenance Cost Rate (Per Mile)

REFERENCES

1. Research Analysis Corporation, "Operation, Maintenance, and Lifetimes of M60 Tanks, M113 Armored Personnel Carriers, and M88 Tank Recovery Vehicles (U)," RAC-T-460, Volume I, February 1965, CONFIDENTIAL; Volume II, May 1965, UNCLASSIFIED.

2. _____, "Operation, Maintenance, and Lifetimes of M60 Tanks, M113 Armored Personnel Carriers, and M88 Tank Recovery Vehicles (U)," Volume II, page 13, UNCLASSIFIED.

3. Operations Research Office, "A System for Inventory Evaluation and Reporting (U)," ORO-T-364, June 1958, CONFIDENTIAL.

A TECHNIQUE FOR DETERMINING EQUIPMENT LIFETIME

by Mr. Howard A. Markham

Research Analysis Corporation

Cognizant Agency: US Army Materiel Command

ACKNOWLEDGMENTS

The author became associated with the work RAC has done on equipment lifetimes at a relatively late stage; the lifetime technique had already been formulated mathematically and made computationally tractable, and most of the applications had been made. That ORO/RAC took interest in the first place is due to the perceptions and energy of Tom Scriggins and Conway Christianson, and the Army Support of Brigadier General William D. Latta, Mr. Maurice Finn and the Deputy Chief of Staff for Logistics in the late '50, Gen Carter Magruder. John Moss made the first formulation of a lifetime equation involving considerations of costs, effectiveness, and obsolescence. Gerald Cooper then considerably revised the lifetime equation and is the author of both its present general form and its specific, computationally tractable form. The overhaul analysis made for trucks is Cooper's; that for tracked vehicles is Norman Agin's. The procedure for assessing the rate of technological obsolescence is due to Boyd Ladd, Margaret Neufer, Bud Smeak, and Harry Sheets. Scriggins and Christianson have overseen all of this work--together until 1962, Christianson alone since then. The exposition generally and the discussions of benefits of the cost amortization approach and the lack of cost discounting are those of the author.

INTRODUCTION

In the past seven years a technique for determining equipment lifetimes has been developed at the Research Analysis Corportion (RAC) and its predecessor the Operations Research Office (ORO). The technique has been used at RAC to determine lifetimes of major Army tracked and wheeled vehicles. The work was done under the primary sponsorship of the maintenance elements of the Deputy Chief of Staff for Logistics and the Army Materiel Command. It has resulted in the series of ORO and RAC

publications numbered one through eight in the bibliography. The findings published in these documents have exerted significant influence on the procurement and management policies of the vehicle fleets studied.

To date the lifetime technique developed at RAC has received no formal exposition except in abbreviated, technical form in appendixes to references 3-8. However, it is the only lifetime analysis known to the author that has had its results accepted by one of the three services and incorporated into the management policies of the equipment fleet analyzed. The present symposium on life-cycle management, then, in addition to seeming a meeting to which a discussion centered about a lifetime technique could contribute ideas of relevance and interest, seemed a good occasion for giving wider exposure to a technique whose results have received recognition but whose ways have gone relatively unnoticed and unknown.

The purpose of this paper is to describe the lifetime technique developed and used at RAC and the general results obtained. The description is set in the specific contexts of materiel life-cycle management on one hand and some general problems of determining materiel lifetime on another.

LIFETIMES AND LIFE-CYCLE MANAGEMENT

Life-Cycle Management

The ultimate aim of life-cycle management of materiel is probably to provide for the adequate performance of a necessary operation while minimizing development, acquisition, support, and disposal cost of required materiel. Figure 1 shows major stages of the materiel life-cycle that must be controlled.

At some time technology has certain capabilities characteristic of the general state of development of the science and engineering communities at large. This is represented by the area labeled "R" for research.

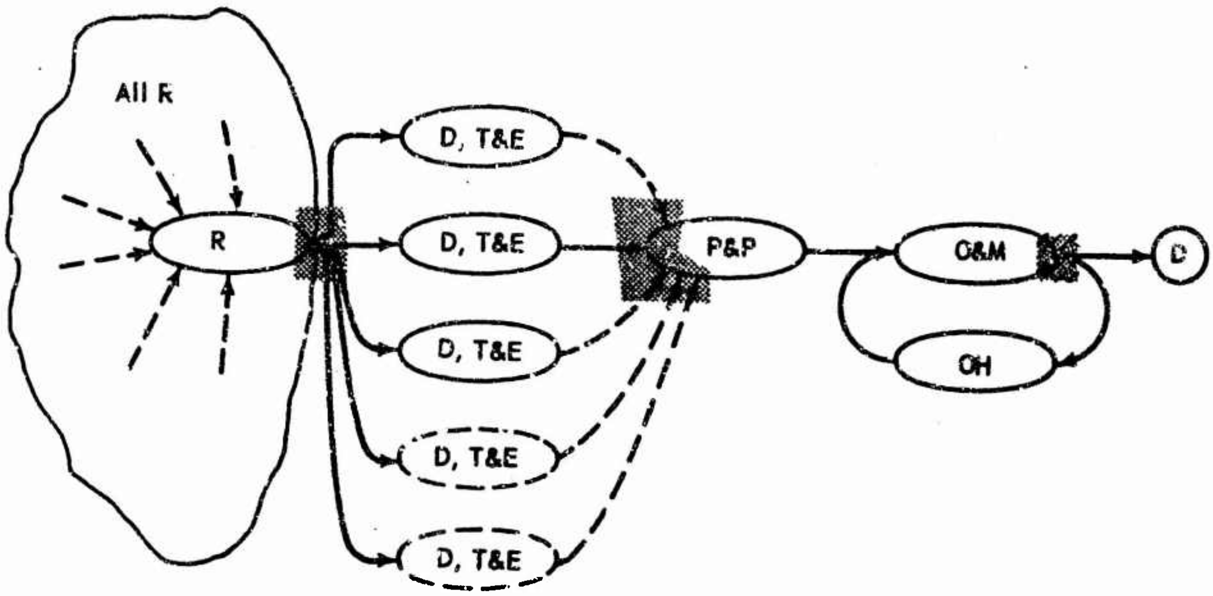


Figure 1 **Materiel Life-Cycle**

////. Area treated by RAC lifetime technique.

Critical decisions in life cycle mgt how many, how often, when?

Given the need for materiel to perform some necessary function, there are a number of configurations this materiel may take that perform the function reasonably well and that reasonably incorporate technological capabilities for producing durability and efficiency. These several feasible configurations are represented by boxes labeled "D, T & E"--develop, test and evaluate--in Figure 1. The purpose of the D, T & E stage is to establish more firmly how the several feasible designs differ so one can choose more intelligently from among them which should be produced and put into operation.

The next two boxes in Figure 1, labeled "P & P"--procurement and production--and "O & M"--operation and maintenance--show these latter stages, and that they represent a selection from among the several possible designs. When operations by a given design can no longer be performed adequately or can no longer be permitted economically, the materiel is disposed of, represented by the last box on the chart, labeled "D".

One other box--under operations--is shown, labeled "OH." It represents the possibility that materiel whose performance has become inadequate or unecomonical can be restored to an acceptable condition by overhauling it. It is really just a way of extending the operating time of the equipment. But it is shown separately because it often requires a major expenditure of time and money, and because it is an alternative to disposal and hence to procurement of a new replacement item.

Optimal management of the life-cycle depends heavily on how funds and effort are allocated among the cycle stages. This allocation is, in turn, critically dependent on decisions made at the points shown shaded in Fig. 1. And to make good decisions at these points requires as a minimum an appreciation of the dynamic environment in which the life-cycle lies.

For example, the first shaded area represents the decisions about how many feasible configurations are to be subjected to D, T & E at the same time, i.e. how many are necessary to reflect adequately the useful variety of approaches afforded by the technology, and how often new configurations should be introduced and old ones discounted. These decisions depend

on one's assessment of the rate and nature of advances in technology on the one hand versus how he supposes they can affect the performance of the materiel of interest and how much this is worth compared with what it will cost on the other. The dynamic elements here are three: The rate of advance of technology, the cost of incorporating the advances that occur over any given time, and the value of effecting any given improvement in the performance of the materiel.

The second decision area represents the choices of when to introduce a new model to the fleet, and which of several alternative new models. This depends on how much better the alternatives are than the model currently in production and what this is worth in the mission environment of the materiel versus how much it costs to introduce the new model. Each element named in this situation is dynamic--it changes over time.

Other critical decision areas shown in Fig. 1 are how many items to produce at what rate, how many to operate for how long, and whether, when, and how often to rebuild. All of these questions must be answered in some way or other in the context of a technology that can produce models of improved costs and/or performance at a rate significant with respect to the aging rate of equipment in use, a mission environment in which the relative value of many mission functions is in a state of flux, and an operating environment in which the logistical system is undergoing frequent changes with attendant changes in the costs of rendering any given support.

Lifetime

In this paper the phrase "equipment lifetime" shall mean the time an item of equipment spends in the box in Fig. 1 labeled "O & M." Equipment lifetime depends on nothing less than all the matters discussed above. Whether an item should be discontinued from operation depends precisely on the cost and performance advantages and disadvantages of substituting any of the available alternative items for continued operation of the current one. And these advantages and disadvantages really depend on the rate of technological advance, the rate of feasible incorporation of advances into test items, and the value of introducing given equipment changes in the prevailing mission environment.

As the cross-hatching of the shading at the end of the operating stage in Fig. 1 shows, the lifetime technique discussed below is a way of answering only one or two of the several critical questions for decision identified for the life-cycle as a whole. Furthermore, although it has been said above that lifetime ultimately depends on decisions taken at the other points in the life-cycle, the RAC technique does not consider such points explicitly. Rather, the technique answers the question: At what equipment age should an item in operation be replaced if the available replacement item has specified acquisition cost, has the same age-dependent operating costs and performance as the current item, but has had the obsolescence incurred by the current item removed? (There are some difficulties with the last characteristic in conjunction with the one preceding it, but further discussion of this point is deferred until a discussion of limitations near the end of the paper.) Thus the explicitly included considerations are acquisition cost of the replacement item, age-dependent costs and performance of operating items, and the obsolescence rate of the current design. Considerations implicitly assumed constant are mission environment in which changes in performance are to be evaluated, logistical system performance, and effectiveness with which incorporation of technological advances in new models and introduction of new models to production is managed.

The lifetime technique being discussed has not tried to answer the lifetime question in the large context of the life-cycle; it has offered an answer to the smaller question of when an item that ages should be replaced by a new, similar item. In doing so, however, the technique gives an idea of the true answer to a major life-cycle management question which has implications for answers to the other major life-cycle questions raised. Two by-products of life-cycle interest associated with the lifetime technique have been partial answers to the rebuild question and an assessment of the rate of technological progress.

THE LIFETIME TECHNIQUE

Characteristics

The lifetime technique being discussed has the following characteristics:

1. The lifetime is defined to be the age at which the average cost per unit of effectiveness is smallest, considering all time since the equipment was issued new. The average cost per effectiveness is derived by summing the cost/effectiveness ratios of each time period of life and dividing by the number of periods. (This is in contrast to summing all the costs and dividing by the summed effectiveness.) It has the algebraic appearance

$$\left(\frac{\overline{C}}{E}\right) = \frac{1}{T} \sum_{i=1}^{T} \frac{C_i}{E_i}$$

for a lifetime T, where C_i and E_i are the costs and effectiveness respectively of time period i.

2. The cost charged to the effectiveness of time period i is the sum of the portions of acquisition cost and maintenance costs amortized in that period. Costs are amortized in equal increments in each period between the time they are incurred and the end of the life being considered. Thus, if the life being considered is T, the acquisition cost is I, and the operation and maintenance costs incurred in the time period j and M_j, the costs amortized in period i are given by the formula

$$C_i = \frac{I}{T} + \sum_{j=1}^{i} \frac{M_j}{T-j+1} .$$

Figure2 shows for a simple case how the amortization procedure redistributes maintenance costs in age. For a lifetime of three time periods, in each of which the maintenance cost incurred is six units, the amortization procedure causes one-third of the six units of cost incurred in the first time period to be charged to each of the three periods of life; it causes one-half of the six units incurred in the second interval to be charged to the two remaining periods of life; it causes all six units of cost incurred in the last time period to be charged to that period. The result

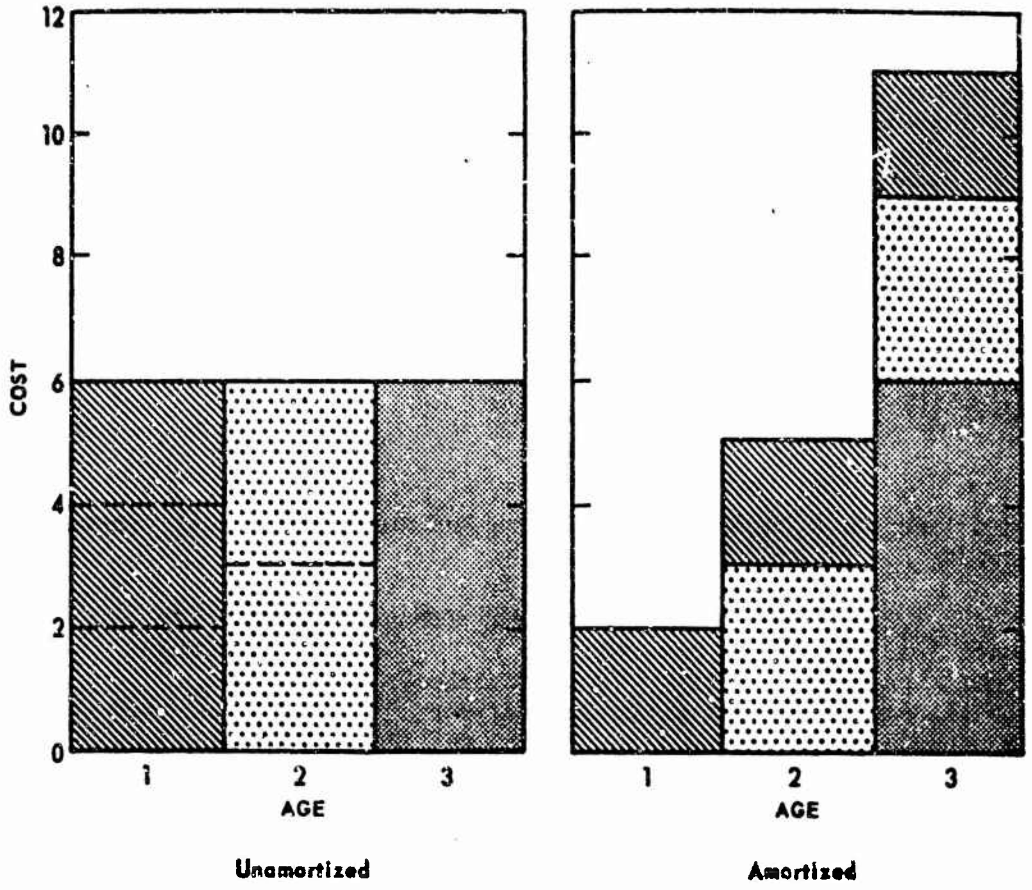


Fig. 2 Effect of Amortization on Distribution of Maintenance Costs in Equipment Age

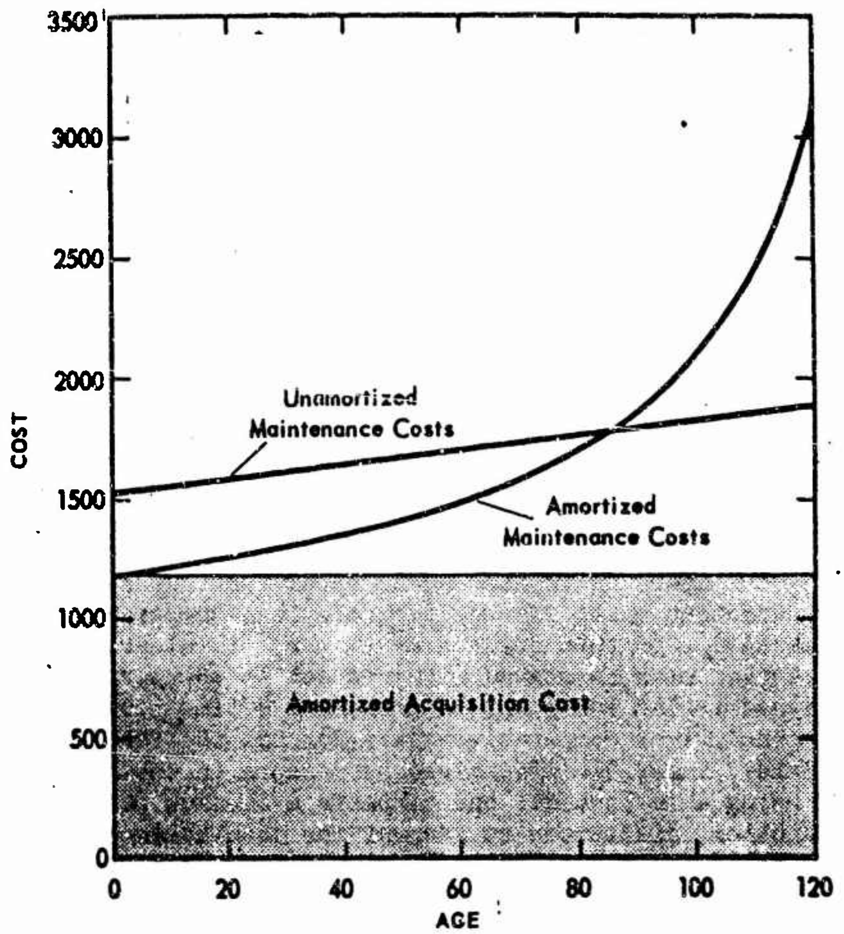


Fig. 3 Redistribution of Maintenance Costs in Equipment Age Caused by Amortization — An Actual Case

is that the distribution of six cost units in each of the three periods of life is transformed to two units in the first period, five in the second, and eleven in the third. An actual case is shown in Fig. 3. Here the amortization of acquisition cost is also shown.

3. Effectiveness is defined as the product of reliability, availability, and an index of non-obsolescence called the technological competetiveness index. Thus,

$$E = (\text{reliability})\,(\text{availability})\,\left(\begin{array}{c}\text{technological}\\ \text{competetiveness}\end{array}\right)$$

Reliability has the usual definition of the probability that a mission of specified content and duration in a specified environment can be completed once begun. Availability also has the usual definition of the probability that an equipment is operable. Obsolescence is assumed to occur continuously in time at a constant rate, r. The technological competetiveness index at some time t is then defined as $(1 + r)^{-t}$. Clearly this is a relative measure stated with respect to the perfect value 1 assigned to the equipment an arbitrary t time periods ago. The technological competetiveness index has the form of a discount rate. Thus, the effectiveness measure is the probability of being able to begin and complete a mission of specified duration, discounted to account for obsolescence. It may be written more compactly as

$$E = \left(\begin{array}{c}\text{probability of}\\ \text{mission success}\end{array}\right)\left(\begin{array}{c}\text{obsolescence}\\ \text{discount}\\ \text{factor}\end{array}\right).$$

Effectiveness assumes values between zero and one, and for any real equipment has a value less than one. A unit of effectiveness is a unit of time at effectiveness index unity, or two units of time at effectiveness $\frac{1}{2}$, etc.

Data Requirements

The data required as input to the calculation are essentially maintenance and cost data; they are summarized on Table 1. The maintenance data required should be basically a list of the breakdowns each vehicle in the sample experienced over some period since issue of the vehicle sufficiently long that trends can be identified and extrapolated into the range of lifetime age with acceptable accuracy. For each

breakdown listed should be known the vehicle age at which the breakdown occurred, the period of time required to restore the vehicle to operability, the man-hours consumed, and the parts consumed. The costs required are the costs of parts at the point of use (i.e. to include their distribution and storage costs), the cost of man-hours, and the cost of a new vehicle.

TABLE 1

DATA REQUIREMENTS

Maintenance Data	Cost Data
Breakdowns, showing	
Vehicle Age	Price of New Vehicle
Time to Repair	Cost of Man-Hours
Man-Hours Consumed	Cost of Parts, to
Parts Consumed	Include Storage and Distribution

One element of required data that does not fall within the two categories named is the obsolescence rate. Obsolescence rate has been determined by administering a pair comparison test to a broad representation of Army interest in vehicles. In this test,[4,5] vehicle designers, operators, commanders, testers, and maintenance men were given a set of cards on each of which a vehicle with particular characteristics--for example weight, size, speed, maintenance requirements--was described. Three of the vehicles described were a past model, the current model, and R & D's best estimate of the next model. The other models were hypothetical. Participants in the test were requested to state which card they preferred of each possible pair of cards that could be formed from the ten. These preference statements were then analyzed to show relative preferences for the three real models. Assigned to an appropriate scale, these preferences were presumed to show the rate of technological improvement as it manifests itself in finished products to the perceptions of users, designers, testers, and maintainers.

Transformation of Data to Lifetime Inputs

Given a breakdown rate $\beta(t)$ at vehicle age t, reliability for a mission of duration d is, for relatively small d

$$R(t) = e^{-\beta(t)\cdot d} \quad .$$

The measure of availability used, the availability potential, may be computed from the data for a vehicle of age t by the formula

$$\beta(t) = \frac{1}{1+\beta(t)\cdot\lambda}$$

where again $\beta(t)$ is the breakdown rate, and λ is the average time out of service per breakdown. The availability potential is the availability that eventually will obtain after the current breakdown rate and average time out of service per breakdown have been operating a long time. It differs from the true availability by an amount dependent on the rate at which β and λ are changing. For the equipment studied it has proved reasonable to take λ as constant and β as changing relatively slowly, so that the difference between availability potential and availability has been small.

Maintenance costs may be computed in straightforward fashion by associating the cost of man-hours with the man-hours consumed and the cost of parts with the cost of parts consumed.

The Lifetime Equation

The definitions made above can be used now to construct a lifetime equation. The average cost effectiveness ratio has been defined in terms which lead to the following expression

$$ACE(T) = \frac{1}{T} \sum_{i=1}^{T} \frac{\frac{I}{T} + \sum_{j=1}^{i} \frac{M_j}{T-j+1}}{E_i}$$

If the number of intervals is allowed to become indefinitely large while the lifetime T is held constant, this expression becomes continuous and can be stated in the form

$$ACE(T) = \frac{1}{T}\left\{ \frac{I}{T} \int_0^T \frac{ds}{E(s)} + \int_0^T \frac{M(s)}{T-s} \int_s^T \frac{dr}{E(r)} \, ds \right\}$$

In the event that operating and maintenance costs M(s) may be approximated by a straight line, say A + Bs, and effectiveness may be approximated by the exponential fe^{-gs}, then this expression may be integrated to yield the closed algebraic form

$$ACE(T) = \frac{I(e^{gT}-1)}{fgT^2} - \frac{A}{f}\frac{e^{gT}}{gT}\sum_{m=1}^{\infty}\frac{(-gT)^m}{m \cdot m!} - \frac{BT}{f}\frac{e^{gT}}{gT}\sum_{m=1}^{\infty}\frac{(-gT)^m}{m[(m+1)!]}$$

By forming tables of the exponential and factorial functions in gT, the solution of this equation for given I, f, g, A, B, and T is quite tractable manually. Four or five trials usually suffice to determine that T which minimizes the average cost effectiveness. This T is the defined lifetime of the item of equipment.

RESULTS

Lifetime

The main elements of this lifetime analysis were developed in the period 1959-1961. Since that time it has provided a consistent approach to measuring average lifetimes of tanks, armored personnel carriers, SP artillery, and 1/4-, 3/4-, 2 1/2-, and 5-ton trucks. The lifetimes derived have seemed reasonable to the Army. They have shown wheeled and tracked vehicles to be two families within each of which lifetimes are essentially similar from vehicle to vehicle. For tracks the lifetime is from 4 to 6 years or 4000 to 6000 miles. For wheels the lifetime is from 7-10 years or from 40,000 to 60,000 miles. These results are presented on Table 2.

TABLE 2

LIFETIMES[a]

Vehicle Type	Lifetime	
	Years	Miles
Trucks	7-10	40,000-60,000
Tracked Vehicles	4-6	4,000-6,000

[a]At observed use rates

Overhaul

With the advent of lifetimes optimum with respect to a single consistent cost-effectiveness criterion it was possible with similar consistency to construct maximum expenditure limits for overhaul or lesser repair as a function of unoverhauled-vehicle age. When one has measured the performance and maintenance costs of overhauled vehicles as a function of the time since overhaul, and has already measured the lifetime of non-overhauled vehicles, then he can set himself the problem: What is the maximum overhaul expenditure for which the best cost-effectiveness after overhaul is as good as the best without overhaul?

In the study on trucks maximum one-time repair expenditure limits, shown in Table 3, were derived as a function of truck age.

TABLE 3

ONE-TIME REPAIR EXPENDITURE LIMITS FOR TRUCKS

Age Interval (Years)	Expenditure Limit (% Of Acquisition Cost)
1-4	60
5-7	25
8-10	10

Source: Reference 3

The policy connected with these limits was that if a repair was estimated to cost more than the limit, the vehicle should be washed out; if the repair were to cost less, the vehicle should be repaired. These limits were stated for three ranges of truck life. They were approximately 60% of the initial cost of the truck during its first four years of life, 25% during its next three years, and anytime an engine required replacement during the last three years. These repair limits were consistent with the computed lifetime of 10 years and measured average performance and costs of overhauled and non-overhauled vehicles.

In the study on tracked vehicles a somewhat different pproach was taken. The overhaul that was then being performed was evaluated and found to yield vehicles with a lifetime about 1/3 that of a new vehicle at slightly reduced average monthly cost and average performance. Maximum expenditure limits were then derived for the situation in which a vehicle was to be retained for a period equal to the sum of the lifetimes of an unoverhauled and an overhauled vehicle. The limits assured that, if the cost of overhaul at some age did not exceed the limit set for that age, the cost effectiveness over the entire retention period for the case when overhaul was effected at that age would not be worse than if no overhaul were made. Of course if the cost of overhaul were less than the limit, the cost effectiveness over the retention period when overhaul is done would be better than if there were no overhaul. Here the results, shown in Fig. 4, were less homogeneous than in the truck case.

Although in neither instance was anything like a comprehensive solution to the question when and to what degree to overhaul given, in both instances answers were given consistent in some way with actually measured performance and costs of overhauled and non-overhauled vehicles and with the lifetimes these performances and costs implied through the lifetime model already described.

One other by-product of these lifetime analyses worth singling out is the measurement of performances and costs. The mere telling of the magnitudes of breakdown rates, downtimes, and maintenance costs for particular models of vehicles in field use constituted pioneering reporting of important information to Army fleet managers and planners.

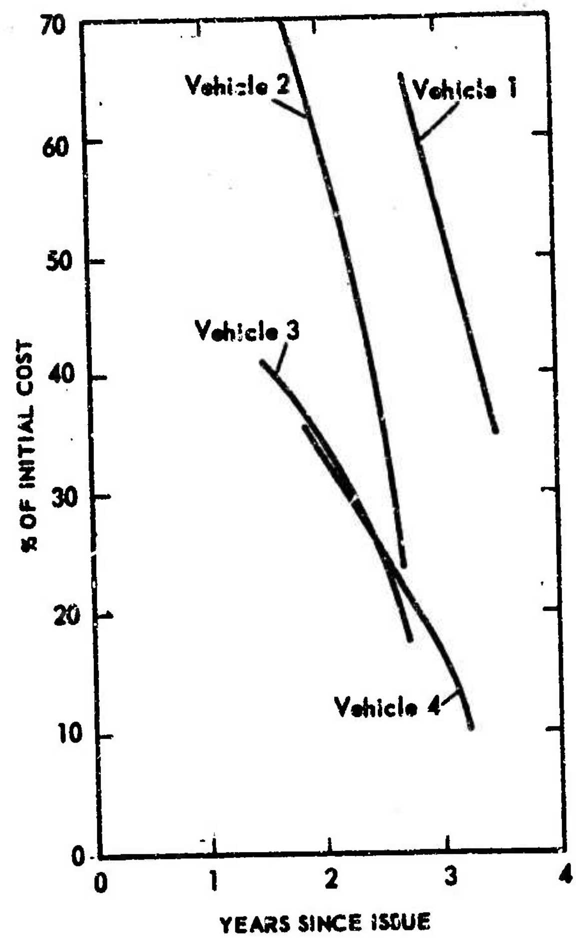


Fig. 4 **Overhaul Expenditure Limits for Tracked Vehicles**

DISCUSSION

Lifetime Criterion - A Problem for the Military

Techniques for determining equipment lifetimes have found widest application in industry, for one major reason: In industry costs and benefits are commensurable. Their difference is profit and a lifetime is defined to be that time at which replacement of current equipment with some alternative equipment will yield greater profit over some specified period of time.

Ideally the military would proceed similarly. Alternative systems and alternative replacement times would be evaluated in terms of their net contribution to the national or world welfare. Only systems promising to make net contribution would be put into operation at all and the life-cycles of inservice systems would be managed in a way that maximized their net contribution. To choose and manage military systems in these terms requires not only commensurability between cost and effectiveness across systems, but also commensurability of various costs within the cost domain of a single system, and similarly for effectiveness. Even these latter do not obtain in the military environment. (They do not in the business world either, strictly speaking, but that is outside the immediate interest of this paper.)

Thus the lifetime analysis discussed in this paper starts at a more modest level than that of the above remarks. Its concern begins after it has already been decided that the system or equipment can make a positive contribution to a relevant welfare and can contribute more than competetive systems or equipments. At this stage some large questions of commensurability have already been by-passed. The question now asked is, "Given that this equipment ages, when should a new item be substituted for a current item?" However, the incommensurability problem remains. Associated with aging are deteriorating performance and increasing maintenance costs. These are incommensurable. Even at this stage of choice the problem is not simply to maximize a profit.

So far as the author knows nobody has seen his way either (a) to suggest how to translate costs and benefits of military systems into commensurable terms, (b) to suggest some goal other than the maximization of the net good, or (c) to suggest how the net good can be determined from incommensurable constituents. In short, so far as is known the dilemma just posed has not been solved except to say that at the point where the incommensurables have been identified and quantified to a reasonable full extent, it is the decision maker who must bring these incommensurables together in his own mind, in the context of his own experience, knowledge, and values, and choose.

Yet, the lifetime technique being discussed yields not a set of incommensurables for a decision maker to mull over, but one number--the lifetime. It makes the decision. But this implies that it has somehow resolved the dilemma just described.'

In fact, it has resolved the dilemma in the case of the Army vehicles studied. Apparently the result of combining costs with the particular measure of effectiveness used in the particular way the lifetime technique requires is a lifetime of intuitive acceptability to the Army--that is, a lifetime consonant with what their perceptions and impressions of the facts of vehicle life gathered through various long and direct experience with vehicles allows them to think of as reasonable.

However, it should be made clear at once that the technique was not devised to resolve the dilemma posed explicitly. Nor has the technique resolved it in any absolute, final, or even authoritative fashion. It is simply that, because the results seem reasonable, and because any lifetime result implies a relation of the kind "effectiveness level x is worth no more than cost y," it seems appropriate to say that the lifetime technique contains an implicit successful resolution of the commensurability dilemma for the equipment studied.

Of course it is not entirely luck that the procedure gives reasonable results. Intuition was at work in its development. For example, if one cannot measure "profit," it is not unreasonable to seek to insure minimum cost for given effectiveness by minimizing the ratio of cost to effectiveness. Nor is it unreasonable to measure vehicle effectiveness in terms of the probability that the vehicle will perform its intended function for a specified time. The next few paragraphs discuss in detail some aspects of the technique that may help to explain why the results are apparently reasonable.

Details of RAC Treatment

There are two ways computation of the cost effectiveness ratio by the RAC technique differs from the gross computation that cause this weighting. One is that in the RAC procedure a reciprocal effectiveness value for each period is computed and then these are combined; in the gross procedure the effectiveness is summed first and then the reciprocal is formed. The result is that periods of high effectiveness contribute relatively less and periods of low effectiveness contribute relatively more to raising the measure of reciprocal effectiveness used in the RAC technique than that used in the gross. Since it is the goal of both techniques to find the lowest cost-effectiveness ratio, for the same set of cost and effectiveness data and for the same treatment of costs, this difference between the two techniques will result in the RAC procedure finding a minimum cost-effectiveness ratio representing a greater proportion of effectiveness units accumulated at higher levels.

The second difference is that the RAC technique amortizes costs over only the life that comes after they are incurred and forms the ratio of the amount amortized in each time period to the effectiveness of that period before it then sums these and takes their average. This is in contrast to summing costs first and then dividing by summed effectiveness in the gross approach. Remembering from Figs. 2 and 3 the shift of maintenance costs to later life caused by the amortization procedure used, it is relatively easy now to say how this second difference also results in preference for effectiveness units accumulated at high levels of effectiveness. The tendency of the RAC treatment of effectiveness to contribute relatively less to increasing reciprocal effectiveness when effectiveness is high is reinforced by making the top half of

the ratio--the cost--smaller in early life when effectiveness is high and larger in later life when effectiveness is lower.

An algebraic description of the two differences between the RAC technique and the gross technique may be made by considering it the goal of each to minimize a ratio of costs to effectiveness for the time the equipment is retained by replacing it at the appropriate age. The procedure is to compute a cost-effectiveness ratio for several retention times until the time for which the ratio is minimum is found. In the RAC technique the ratio is computed by

$$\left(\frac{\overline{C}}{E}\right) = \frac{1}{n} \sum_{i=1}^{n} \frac{C_{ai}}{E_i} \qquad (1)$$

for a retention time of n units of time. The gross procedure is

$$\left(\frac{\overline{C}}{E}\right) = \frac{\sum_{i=1}^{n} C_i}{\sum_{i=1}^{n} E_i}$$

which is equivalent to

$$\left(\frac{\overline{C}}{E}\right) = \frac{1}{n} \sum_{i=1}^{n} \frac{\overline{C}}{\overline{E}} \qquad (2)$$

The difference between (1) and (2) is that in (2) the same cost and effectiveness are used for each time period, while in (1) different cost and effectiveness are used for each period, and, as the subscript a is meant to designate, the costs used in a time period are those amortized in it.

A point not heretofore raised is that costs are not discounted in the RAC approach. If it is considered proper to discount future costs to account for the reality that future costs demand from current assets amounts less than the cost by the interest that can be earned meantime, then the failure of the RAC technique to do so is a third way that effectiveness units accumulated at high levels are preferred. If discounting were used the costs charged to future periods would be smaller than undiscounted costs the more future they were. But since discounting is not used, and since earlier periods have higher effectiveness than later periods, not discounting results in increasing higher cost numerators in the cost-effectiveness ratios of periods of increasingly bad effectiveness.

Figures 5 and 6 show the three effects described for the set of hypothetical data shown in Table 4.

TABLE 4
HYPOTHETICAL COST AND EFFECTIVENESS DATA

Age Period	Age Dependent Costs		Age Dependent Effectiveness
	Undis-counted	Dis-counted	
1	6	6.00	.9
2	7	6.65	.8
3	8	7.20	.7
4	9	7.65	.6
5	10	8.00	.5
6	11	8.25	.4

Initial cost: 50

In Fig. 5 the cost-effectiveness ratio for various retention times is shown for several ways of computing cost-effectiveness ratios. The bottom curve was computed by the

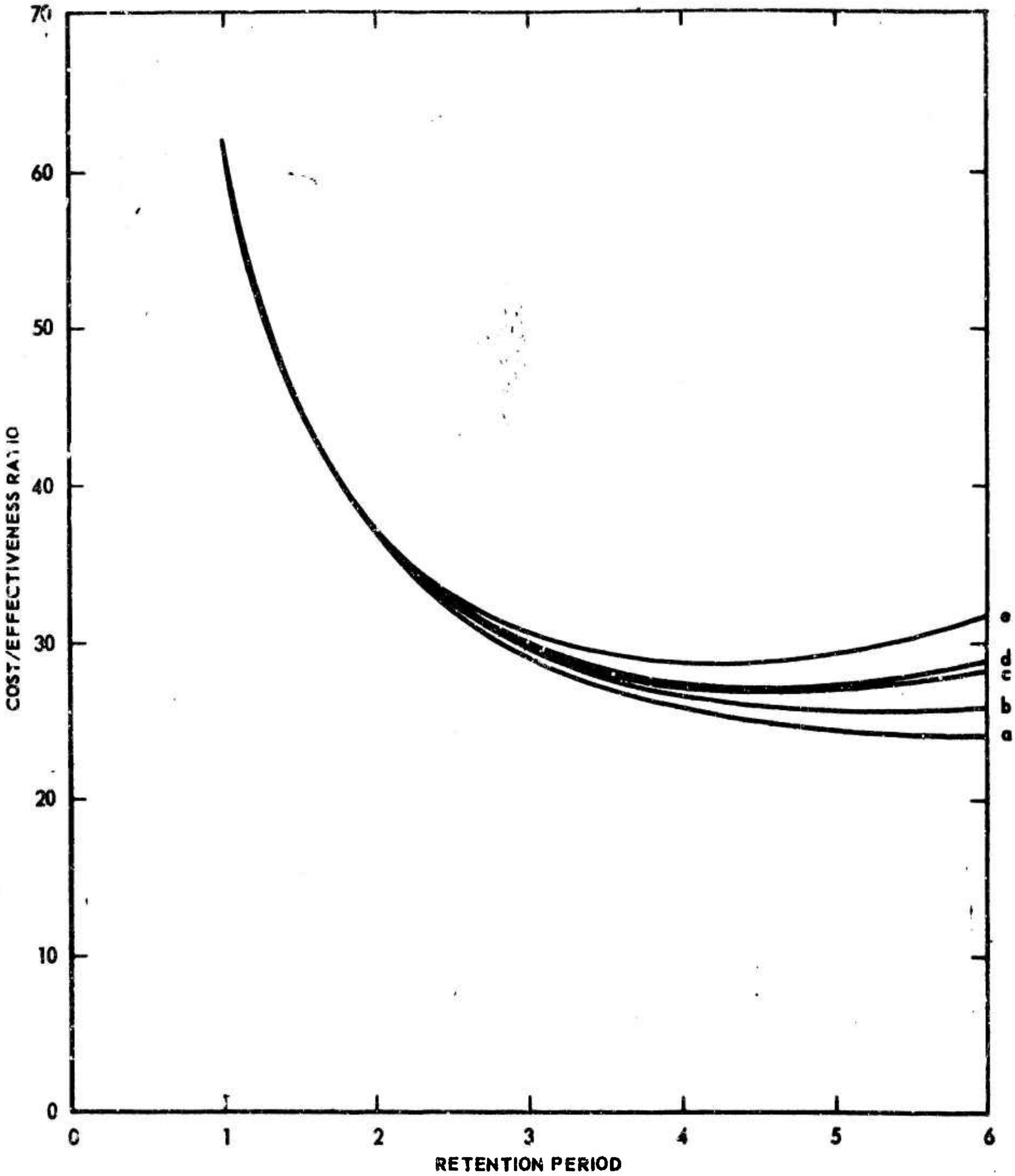


Fig. 5—Cost-Effectiveness Curves for Various Kinds of Cost-Effectiveness Ratios

a. Average discounted cost/average effectiveness.
b. Average undiscounted cost/average effectiveness.
c. Average undiscounted cost × average reciprocal effectiveness.
d. RAC procedure except actual, not amortized, maintenance costs are used.
e. RAC procedure.

gross technique using discounted costs. The discounting used was a simple five percent per period--probably unrealistically simple and high but suitable for illustrative purposes. The second curve up differs only in that costs were not discounted. The third curve differs from the bottom curve in that costs were not discounted and a cost-effectiveness ratio was computed for each period using average cost for the numerator but actual effectiveness for the period in the denominator. In the next curve up all the previous changes are retained while use of the costs actually incurred in the period as the numerator of the cost-effectiveness ratio of the period is added. The top curve of course is the RAC curve, different from the previous one only in the use of amortized costs in place of actual costs. Each curve indicates a shorter life and hence a life consisting of relatively more time spent at high effectiveness. Thus each change in the computations leading to the curves may be regarded as introducing preference for high effectiveness.

In Fig. 6 a different set of curves is shown to represent the effects of the various computations. The point of view here is that, if it is assumed that the gross procedure is the only justifiable way of forming a ratio of costs to effectiveness, then the cost-effectiveness curves of Fig. 5 all represent use of the gross procedure but differ in the value system that expresses how much a given level of high effectiveness is preferred to a given level of lower effectiveness. For example, the lowest curve of Fig. 6 shows how the actual effectiveness curve--the top curve--should be changed to allow the gross technique to give the same cost effectiveness results as the RAC technique. The lowest curve can be interpreted as showing that a measured effectiveness level of 0.70, for example, (top curve) is rated by the RAC technique as being of only 0.61 (bottom curve) value for operational purposes. Thus, the differences between the RAC technique and the gross technique may be given the concrete interpretation of being a way to introduce a value system that translates effectiveness measured in terms of availability, reliability, and obsolescence into the value of this effectiveness to operations.

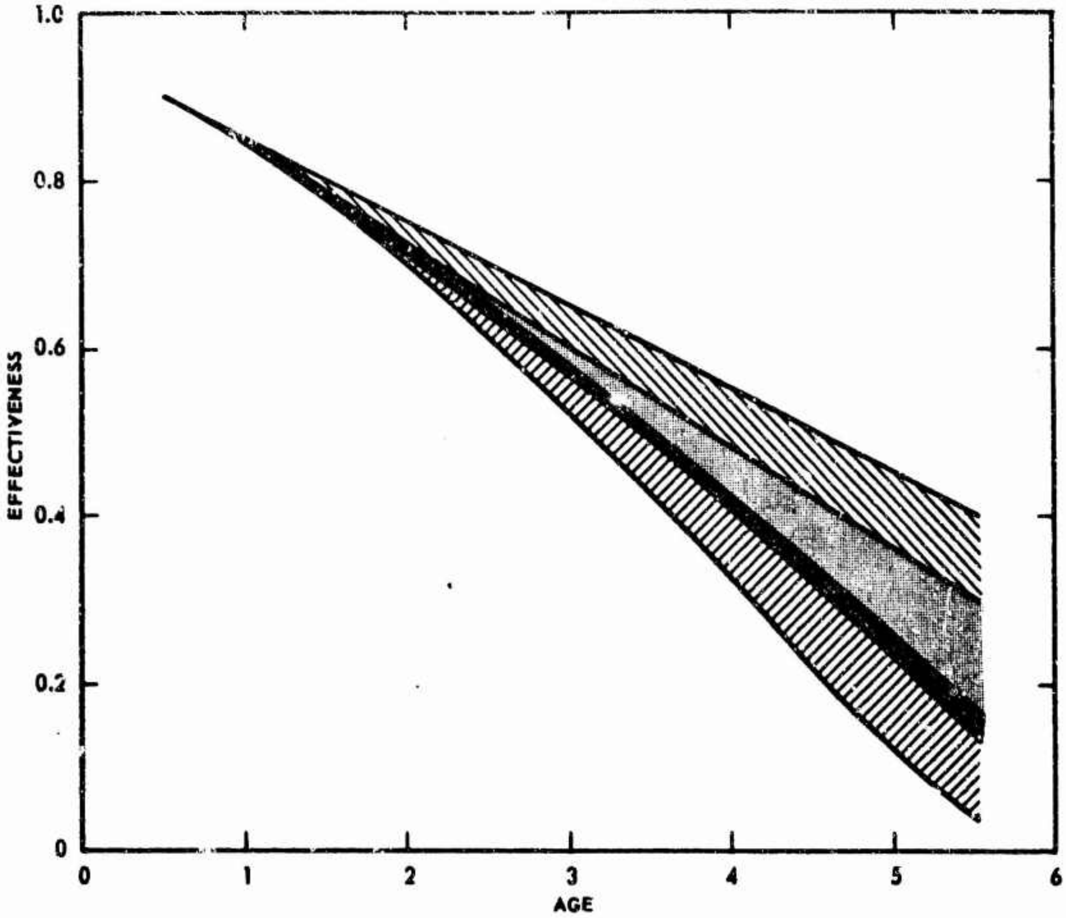


Fig. 6—Incremental Degradations of Effectiveness that Would Yield the Same Cost-Effectiveness Ratios as the Gross Ratio of Total Discounted Costs to Total Effectiveness, for an Accumulation of Various Deviations from the Gross Ratio

- No discounting.
- Average reciprocal effectiveness instead of reciprocal of average effectiveness.
- Maintenance costs charged to intervals incurred in.
- Maintenance costs amortized.

Limitations

A few limitations of the technique may be noted. It computes an average lifetime. Optimum replacement times for individual vehicles vary, of course. Dean reported in 1961[10] that it was characteristic of the lifetime analyses then in being that only an average value was derived; one stochastic model had been developed.

The RAC model in effect assumes an infinite sequence of replacements. In reality one may be concerned with optimizing only the next five or ten years with current decisions. Thus if the RAC model yielded a lifetime of five years and the equipment was needed for only the next eight years, the optimum policy might be to operate the first item four years and the second four years. The model would not automatically say this. Its results really mean five years is the optimum retention time only if the period of interest is very long or is a multiple of five years.

The treatment of obsolescence in replacement vehicles is awkward. The statement that the computed time applies if there are to be a long sequence of replacements is strictly true only to the extent that replacement vehicles have the same acquisition cost and age-dependent costs and effectiveness as the vehicles measured and each replacement begins with zero obsolescence. The idea that obsolescence can be removed while the costs and effectiveness remain the same does not sit well. One might prefer to start the new replacement at the obsolescence level its predecessor left off at. This would seem more realistic. However, it seems likely that it would alter the lifetime of the replacement vehicle very little, and that the main effect would be simply to inflate the magnitude of the cost effectiveness ratio obtained. In any case there is probably some reasonableness in the idea that engineering improvements in a given model partially offset obsolescence while affecting the measured costs and the effectiveness imperceptibly.

Although obsolescence is included explicitly in the calculation, it is used essentially as a way of discounting late model life effectiveness to make a realistic accounting for the loss in competetiveness of a given design. It does not mean that therefore the lifetime derived is the time to change designs. That question is not answered by the analysis discussed. However, the results should be of some help to those who must decide when to introduce a new model.

The technique is concerned only with a kind of "economic" life. The possibility that the tactical value of a tank, for example, may have become unacceptably low after only half of its economic life has expired is not considered. This must be treated separately. Of course such a decision would pre-empt the replacement decision made by the RAC technique. An economic replacement time is applicable only if a more critical replacement time does not precede it.

In conclusion, it should be noted that while it has been stressed and much has been drawn from the fact that the Army has accepted the results as reasonable, this does not mean they have employed them without regard for their limitations. They are aware that the results are average values not rigidly applicable to every individual vehicle; they know that tactical considerations may nullify or modify the relevance of the results, especially in high priority units; they know that the question of model succession is not resolved by the calculations described. In fact, there is probably no basis for assuming that they cannot be persuaded in the future that results derived by the technique described are not reasonable, or at least that some other results are more reasonable.

BIBLIOGRAPHY

1. Operations Research Office (now RAC), "Description of the Model and Computer Program (U)," Appendix C of "A System for Inventory Evaluation and Reporting (U)," ORO-T-364, January 1958. UNCLASSIFIED

2. ____, "Operation, Maintenance, and Cost Experience of ¼-Ton Truck Fleet (U)," ORO-T-382, May 1961. CONFIDENTIAL

3. ____, "Economics of Maintenance and Replacement of 3/4-, 2 1/2-, and 5-Ton Truck Fleets (U)," ORO-T-401, September 1961. CONFIDENTIAL

4. Research Analysis Corporation (was ORO), "Evaluating Improvement in Mission Performance of Military Equipment (A Method of Comparing Effects of Technological Change on Military Item Effectiveness) (U)," RAC-TP-58, May 1962. UNCLASSIFIED

5. ____, "Operation, Maintenance, and Cost Experience of the Tank, Armored Personnel Carrier, and Self-Propelled Howitzer Vehicle Fleets (U)," RAC-T-409, September 1962. SECRET

6. ____, "Economics of Maintenance and Replacement of 5-Ton 6x6 Medium Wrecker Truck M62 (U)," RAC-SP-186, October 1962. CONFIDENTIAL

7. ____, "Operation, Maintenance, and Lifetimes of M60 Tanks, M113 Armored Personnel Carriers, and M88 Recovery Vehicles (U)," RAC-T-460, February 1965. CONFIDENTIAL

8. ____, "Aging and Lifetime of the M151-¼-ton Truck (U)," draft proposed RAC TM.

9. George Terborgh, <u>Dynamic Equipment Policy</u>, McGraw-Hill, New York, 1949.

10. Burton V. Dean, "Replacement Theory," Ch. 8 in <u>Progress in Operations Research</u>, Vol 1, ed. Ackoff, John Wiley & Sons, New York, 1961.

11. Arthur D. Hall, A Methodology for Systems Engineering, Van Nostrand, Princeton, 1962.

ABSTRACT

MEASURES OF EFFECTIVENESS OF ATTACK (MEA)

AUTHOR: 1/Lt Brian Smith
Engineer Strategic Studies Group
Washington, D.C.

The MEA study, produced by the Special Weapons Division, Engineer Strategic Studies Group, is a prototype model of the US industrial and vocational structure as it may emerge after a large-scale nuclear attack. Residual economy and population are analyzed vis-a-vis their combined abilities to recover and sustain an acceptable post-attack level of output. The loss of certain critical skills and their interrelationship to input/output requirements of the industrial sector of the post-attack ~~economy~~ are examined by the use of a computer model.

MEASURES OF EFFECTIVENESS OF ATTACK

Lt. Brian R. Smith
Engineer Strategic Studies Group
Washington, D.C.

In September of 1964, the Engineer Strategic Studies Group was requested by DCSOPS to undertake a study dealing with measuring the effectiveness of strategic attacks, as the current measures were considered to be unsatisfactory. Therefore, a project entitled "Measures of Effectiveness of Attack (MEA)" was begun. Since that time, the basic methods developed in the MEA project have been put to use in a study entitled "Post-Attack Viability of the United States in 1975 (PAVUS-75)".

More specifically, MEA is an investigation into the problem of developing more meaningful measures for evaluating the effectiveness of strategic attacks upon a nation. The principal foci of concern are the post-attack postures of the industrial and demographic sectors of the United States.

Measuring the effectiveness of a strategic attack is currently assumed by many to be a problem of straightforward damage assessment. Results of strategic nuclear exchanges have been reported in terms of population casualties, destruction of manufacturing value added or the loss of industrial floor space. These measures have been further broken into sub-categories or sectors but this breakdown has not

identified the attack effectiveness, because the results have all been based on the principle of equal utility. For example, some population results have underscored the loss of certain professions or occupations. However, no significant techniques have been developed which take into consideration the post-attack utility of these various classes of the population towards the achievement of national recovery goals.

Before a measure of effectiveness can be developed, the desirable attack objectives must be defined. These may consist of one or more of the following:

1. Damage to and denial of industrial facilities including those necessary to support military operations.

2. Casualties and fatalities from blast, fallout, and fire.

3. Damage to capabilities to move military forces to a theater of operations.

4. Destruction of counterforce targets.

Many previous studies have considered only one basic phase in evaluating a strategic attack--physical damage which then is related to the ability of the nation to recover. Figure 1 demonstrates the attack and recovery phases of a strategic attack. The attack phase shown here is climaxed by a vertical line, at which time the total physical damage is determined. The latter phase of the analysis, in this example, indicates that the opponent which was damaged most has in fact been able to recover at a faster rate than his opponent,

reaching a level of decisive victory at an earlier time. Although this earlier recovery by the defender will depend upon many segments of the post-attack environment rising to the task of economic recovery, the nature of input/output relationship within the country in question will be of major importance. In other words, the country which has placed its manufacturing facilities close to its sources of supply will be placed in a more vulnerable attack position than the nation which has a more decentralized structure.

There is a very important step missing in such an analysis as shown here: the actual transition between the damage assessment and the recovery phase which occurs at the point of maximum degradation. Physical damage to the industrial segment of the economy will not be a sufficient measure of the true residual assets; the additional impact of the loss of input (raw materials, supplies) must be considered as well. That is to say, there is a synergistic effect to be considered when evaluating the attack and that the true point of maximum degradation is below the point determined by subtracting economical losses from pre-attack assets. This further reduction in capability may have a decided impact upon the ability of a nation to recover. The model considers the translation from physical damage to recovery assets and determines the true point of maximum degradation. Only when this is accomplished do we have a complete analysis and three points of departure for such a study might be:

1. Indicator Industry Study. Such a study would attempt to determine a specific industry which would experience an output correlating closely with that of the national economy. To coin a phrase, "as the steel industry goes, so goes the nation."

2. Weighted Population Base Study. Previous studies have been adequate in quoting the numbers of casualties, the fatalities by area, cause, and time. These raw numbers, other than for their staggering size, have been of little value except for evaluating the adequacy or shortcomings of the emergency medical procedures in the event of a nuclear war. It is the next step that must be addressed, that of assigning a relative utility value to the various professions and industries. Some measure must be developed to reflect the difference in recovery value to the nation among vital professions.

3. Discrete Area Analysis Study. The discrete area analysis model would involve the determination of those geographical areas considered essential to the stated objectives of an attack. This might include the identification of prescribed key industries, large urban centers, and important military targets. The vulnerability of these elements would then have to be determined and their current and potential value related to the total mobilization effort.

The MEA analysis most closely resembles the third study.

As stated at the beginning, ESSG was asked to find a more exact method of determining post-attack effectiveness of the nation.

Before describing the model in detail, it may be worthwhile to point to a few problem areas in a study of this nature.

1. The vital level of population and industrial bases cannot be accurately assessed.

2. Exact emergency conditions cannot be predicted.

3. Critical assets are not ranked in order of post-attack importance.

4. Industrial recovery data is inadequate.

5. Computable data bases are cumbersome.

MEA has attempted to decrease the exposure to these problems by:

1. Using total shipments of material instead of MVA, thus providing an input/output relationship, and a meaningful measure of the actual economic value of goods and services available to the post-attack environment.

2. Incorporating "critical" economic data.

3. Utilizing an interindustrial relationship.

Figure 2 is a simplified flow chart of the information flow throughout the MEA analysis. Two distinct areas of operation are included as the attack phase, the point at which damage levels are assessed as a result of a particular weapon laydown and the MEA phase in which the damage levels are translated into meaningful urban/industrial losses to the national economic posture.

Data on weapons (HOB, CEP, yield), wind (speed and direction), visibility, facilities (hardness, sheltering), and anti-ballistic missile

defense (if desired) are entered into a damage assessment routine,* the results of which are presented as a measure of the fatalities due to blast and fallout. The percent damage to industrial structures is taken directly from the weapon blast level computed by the attack model; the damage to population is reckoned as the combined fatality rate due to blast plus fallout.

The actual loss to the occupational sector of the economy is simply subtracted from the pre-attack population to determine the residual labor force. This information is then used along with the industrial residual capability to determine if an imbalance exists between industry and labor force and if both sectors are equal to the task of contributing to a viable economy in the post-attack era. Levels of criticality must, of course, be established so that the residual data base may have a basis of comparison. If the factors of labor and capital fall short of their minimum criteria, such pre-attack measures must be taken to assure survivability (namely, the stockage of vital raw materials and machines and the pre-training of less skilled workers in professions that may be vital after a nuclear strike, such as first aid and firefighting).

A final refinement of the post-attack data may include such dependencies as:

---

* specifically, the Damage Assessment of Hazard (DASH) model developed by the System Sciences Corporation for the use of the Office of Civil Defense.

1. Industry to industry (completed in MEA).

2. Industry to population (food, clothing, medical).

3. Population to industry (labor force).

4. Population to population (doctors).

A brief explanation of the method for analyzing the industrial interrelationship as contained in the MEA study follows.

In order to utilize this input-output relationship, specific input is required. The input requirements for this model are the actual value of shipments per industry per target area and their actual relationship as to an overall customer/supplier standpoint.

It would be relatively impossible to predict the exact effect any particular industry would have upon another, and almost insuperably difficult, without extensive research, to predict the exact dependency of the industry upon the labor force and the exact relationship between different trades in the population base.

The MEA analysis attempts to circumvent this situation by presenting three effects that any industry will undergo as a result of the attack:

1. Immediate physical effects to industrial production capability,

2. Maximum interdependent effects, and

3. Minimum interdependent effects.

Thus, it was stated that the immediate effects, those caused by the blast, are included in the program results. These give a measure

of the effect upon an industry due to direct damage, and th , is the system that many measures of effects studies utilize. Certain damage results will have a great impact upon the industrial structure; this impact is given a maximum and a minimum level.

Second, the maximum interdependent damage effects are computed. The theory behind this upper boundary of expected loss assumes a one-for-one relationship between input and output. If the motor industry is destroyed, then the automobile industry will not produce. If the motor industry is 50 percent damaged, then the automobile industry may only produce 50 percent of its volume.

Finally, the lower limit, or the minimum interdependent damage effects, assumes that a firm obtains items in varying quantities and that only at a certain upper boundary will production be fully degraded. Looking again at the automobile industry, if the firms supplying chrome metal are destroyed, the automobiles may still be produced but at a lower cost.

Assume a contained economy with three cities, each with four industrial facilities as shown in Table 1. The value of each industry is shown in millions of dollars. To translate values into meaningful numbers for computation, the total values for each industry are summed and divided by the total giving a percentage or ratio of industry in that area (Table 2).

INDUSTRY

		1	2	3	4
	A	20	10	50	80
City	B	70	40	20	10
	C	10	50	30	10
Total		100	100	100	100

Totals of value of shipments per industry ($ in millions)

Table 1

INDUSTRY

		1	2	3	4
	A	.20	.10	.50	.80
City	B	.70	.40	.20	.10
	C	.10	.50	.30	.10
Total		1.00	1.00	1.00	1.00

Ratios of value of shipments per city

Table 2

Table 3 gives the actual input/output relationships between various industries. Reading down column one, industry 1 requires none of its own output, 10 percent of 2, 10 percent of 3, and 80 percent of 4. Also, since we wish to reprecent the maximum effect that can take place, we shall create Table 4 which shows that, if any interrelationship exists, a "1" will be placed in the new matrix; if no relationship exists, a "0" will be placed.

INDUSTRY

		1	2	3	4
	1	0	.8	.1	0
	2	.1	0	.8	.7
Industry	3	.1	0	0	.1
	4	.8	.1	0	0

Actual input factors

Table 3

INDUSTRY

		1	2	3	4
	1	0	1	1	0
	2	1	0	1	1
Industry	3	1	0	0	1
	4	1	1	0	0

Maximum-Minimum input factors

Table 4

Table 5 shows the summary calculations for Industry 1 in City A, immediate damage is 20 percent of Industry 1 which is contained in City A; maximum expected effect is the 80 percent of Industry 4 which was damaged and now cannot supply Industry 1 on a one-for-one basis. The minimum effect is the actual amount of Industry 4 (80 percent) in City A times the interindustrial factor between 1 and 4 (80 percent), or 64 percent.

<u>IMMEDIATE</u>

20% of Industry 1

<u>MAXIMUM</u>

80% of Industry 4 is lost to Industry 1

<u>MINIMUM</u>

64% of Industry 4

(80% in City A x 80% interindustrial factor)

Table 5

Summary Calculations, City A, Industry 1

Thus, when the interindustrial factors are assessed, the probable damage to Industry 1 is seen to be considerably higher than if straight blast were used.

It can be seen that with two (or more) interconnected industries, endless "mirror images" of relationships exist. If Industry A loses some fraction of its output to Industry B and Industry A is dependent upon Industry B for input then the "second round" loss to Industry A will include:

1. The actual structural damage to the plant and equipment of Industry A.

2. The loss of input to Industry A as a result of the physical damage to Industry B, upon which Industry A depends for supplies.

3. The increased loss of input to Industry A as a direct result of Industry B not reaching the theoretical limit of its degraded capacity because it did not receive the full quota from Industry A.

This cycling process can continue indefinitely. Actually, the interindustry model considers three iterations as sufficient to establish a point of convergence within the industrial relationships which exist.

The national industrial loss as a result of direct damageand the added effect of the interindustrial input loss may be graphically shown. Figure 3 is an example of the cumulative effects of direct and input effects as the number of counties increase; this graph represents sample data for the crude petroleum and natural gas industry. The lower

line represents the percentage of the total national petroleum and natural gas as contained within the 272 largest countries; the upper line is the combined effect of both the industrial facility itself plus the input of supplies and material necessary to produce the final product. The distance between the upper and lower curves shows the additional percentage loss which must be considered in assessing recovery capability post-attack.

Figure 3 demonstrates the additional effect of the input/output structure for the crude petroleum and natural gas industry (taken as an example). If the 100 largest cities (composed of the 272 largest counties) were destroyed, then the loss to this industry as a result of direct physical damage would be approximately 38 percent of its pre-attack capability. With the damage to suppliers, however, the industry would have a 52 percent capacity available (100 percent minus 48 percent).

When the 78 industries used in the study are classified according to their interdependence upon inputs from other segments of the economy, these dependencies show various levels which may be listed as high, high moderate, moderate, low moderate, and low. In general, as the degree of input dependency decreases, the concentration within the large cities increases. Figure 4 is a scatter diagram of these two variables plotted against each other. The "least squares" equation which describes this diagram is

$$y = 0.28 - 0.23x$$

where y is the degree of input dependency and

x is the percent of industry located within the 100 largest cities

and this equation has a coefficient of correlation of -0.41 for the pertinent data. If the general case holds that the independent industries are located downtown (thus being in a more vulnerable position because of the destruction potential of an urban/industrial target) there must exist some "Achilles' heel" industry or industries. Such industries would be defined as those vital to the recovery of the nation which would at the same time suffer high losses as a result of being in geographically vulnerable locations. It might be assumed that an industry requiring a high degree of input (chemicals, primary iron and steel, crude petroleum) would not suffer a loss as heavy as an independent industry since the changes of all input being severed would be low. From an examination of the study data, it was noted that machine shop products, glass and glass products, transportation equipment, farm machinery and equipment, scientific and controlling instruments, and the optical industries are highly vulnerable to crippling losses in a large-scale attack as well as being vital to a viable economy. Also the motor vehicle and the aircraft industries, although not classified as independent, would be suspect to scrutiny due to their concentration in the urbanized areas.

In our present analysis of the 1975 data base for 272 counties and 78 industries, the county/industry matrix consists of 21,294 data points whereas the two interindustrial tables occupy a combined core space of 13,136 locations. Earlier feasibility studies were run on an IBM 7090, but it was soon discovered that with the necessary space for the program, the loader, and the buffer storage, size limitations would predicate the use of the 65K core size of the CDC 3600. The routine which handles the loss to the occupational work force is small enough for a 16K MH 800. The results of the industrial damage program may be run for each target area which takes approximately 10 minutes of CDC 3600 time and 2 hours of printer time.

In order that the 272 counties considered in the PAVUS-75 study may be ranked in some order of significance, a measure which weighs the effects of industry and population contained within the counties must be developed. Any measure which may be used will be suspect since a "dollar-to-man" equivalent is impossible.* A singular tabulation of industry would ignore the effect of population, and vice versa. <u>In general</u>, however, the population does center itself around the large industrial cities. (Queens, New York, and Kings, New York are examples of counties with no industry but together containing 2 percent of the national population.) The program's target list weighs the 272 counties

---

* An exception is the case of certain unique industries (e.g., construction) where the output of both men and machines may be averaged on a time equivalent, compared, and related to a dollar equivalent.

by population and industry. The values represent the fraction of population, the fraction of industry (including a measure of the interindustrial relationships which exist), and a combined fractional weighting on a one-for-one basis by percentage.

Now that the targets have been ordered, it is possible to see just how much industry and population is contained in the counties when grouped together. Table 6 shows some of the sample calculations for the five largest counties according to the previously described weighting. The column marked "Overall Weighting" is simply the sum of the two preceeding columns totaled for all 272 areas and expressed as a percent of that total. Roughly, these overall weighting values may be used to express relative "worths" of geographic areas when compared to the national picture and to each other.

SAMPLE TARGET AREAS

Rank Order	County Name and State	Fraction of Industry	Fraction of Population	Overall Weighting
1	Los Angeles, Cal	.03773	.04718	.04438
2	Cook, Ill (Chicago)	.01755	.02593	.02343
3	Milwaukee, Wis	.03272	.00530	.01346
4	New York, NY (Manhattan)	.03167	.00550	.01329
5	Cuyahoga, Ohio (Cleveland)	.01341	.00842	.00990

Table 6

With the weighting system being used, it is an easy task to accumulate the percentages of population and industry. Figure 5 is a graph of the cumulative percentages values of both industry and population plotted against the number of counties arranged in descending order of overall weighting. If it is desirable to know what number of counties must be totally destroyed to leave the United States at levels of 50 percent of its pre-attack population or industry for example, the values of 52 and 132 counties may be read, respectively.

There are many steps of refinement that may be undertaken to improve the basic philosophies of the MEA study since the work done so far has been of a probing nature in an attempt to come up with a working methodology. One step in the refinement process would be a rather simple form of expansion or model refinement; i.e., the data base expanded to include more target areas.

Further refinement to the study could introduce such factors as transportation, the time delay necessary to supply industries with critical materials; communication networks which were destroyed in the attack; the loss to utilities and the inherent time to bring these facilities back on the line; and the pre-attack capacities of the industries which would measure what maximum output constraints would bound the upper limit of production.

Secondary objectives of expansion and refinement might be the development of grading factors which would identify the individual

characteristics of target zones; namely, the density of population and industry, their combined effect upon national recovery, and the time-phased interrelationship of all target zones (even to include foreign imports) through interindustrial relationships and the transportation networks.

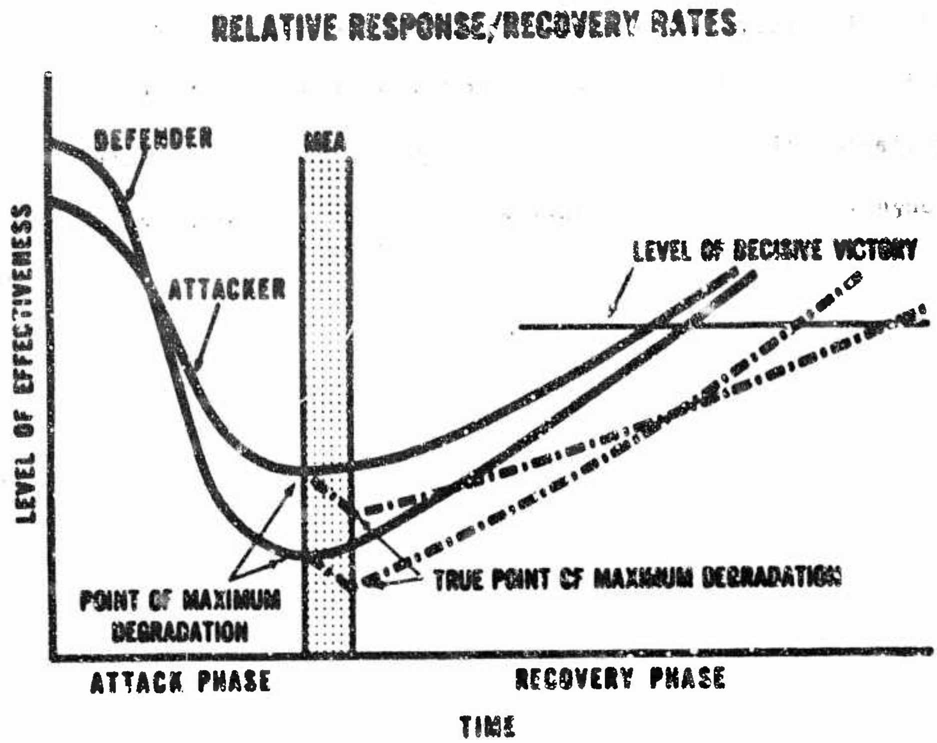

RELATIVE RESPONSE/RECOVERY RATES
DEFENDER
MEA
ATTACKER
LEVEL OF DECISIVE VICTORY
LEVEL OF EFFECTIVENESS
POINT OF MAXIMUM DEGRADATION
TRUE POINT OF MAXIMUM DEGRADATION
ATTACK PHASE
RECOVERY PHASE
TIME

FIGURE I

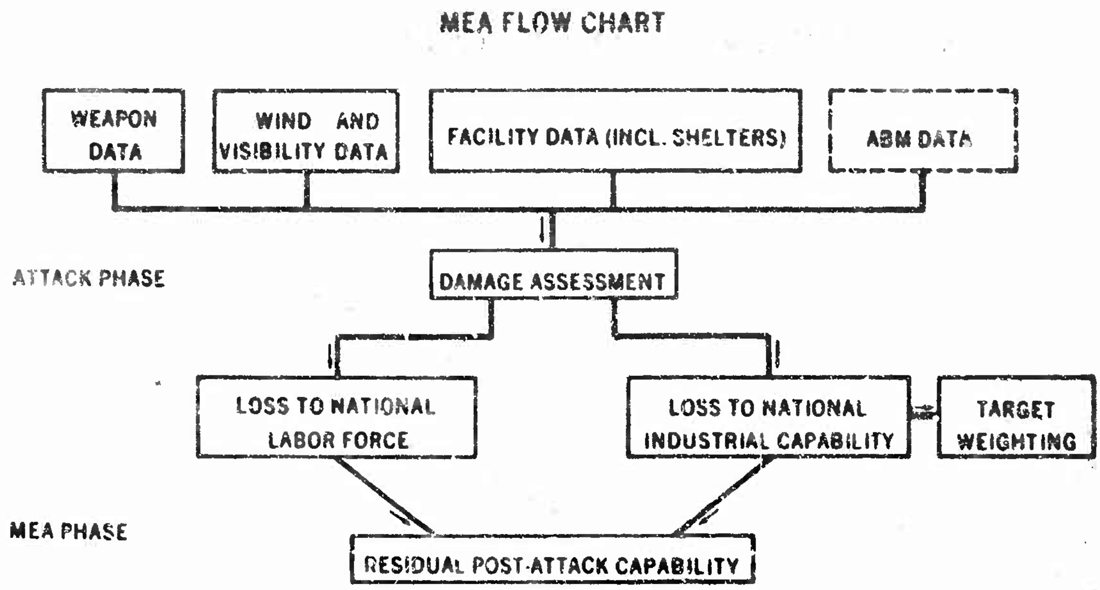

MEA FLOW CHART
WEAPON DATA
WIND AND VISIBILITY DATA
FACILITY DATA (INCL. SHELTERS)
ABM DATA
ATTACK PHASE
DAMAGE ASSESSMENT
LOSS TO NATIONAL LABOR FORCE
LOSS TO NATIONAL INDUSTRIAL CAPABILITY
TARGET WEIGHTING
MEA PHASE
RESIDUAL POST-ATTACK CAPABILITY

FIGURE 2

CRUDE PETROLEUM & NATURAL GAS INDUSTRY

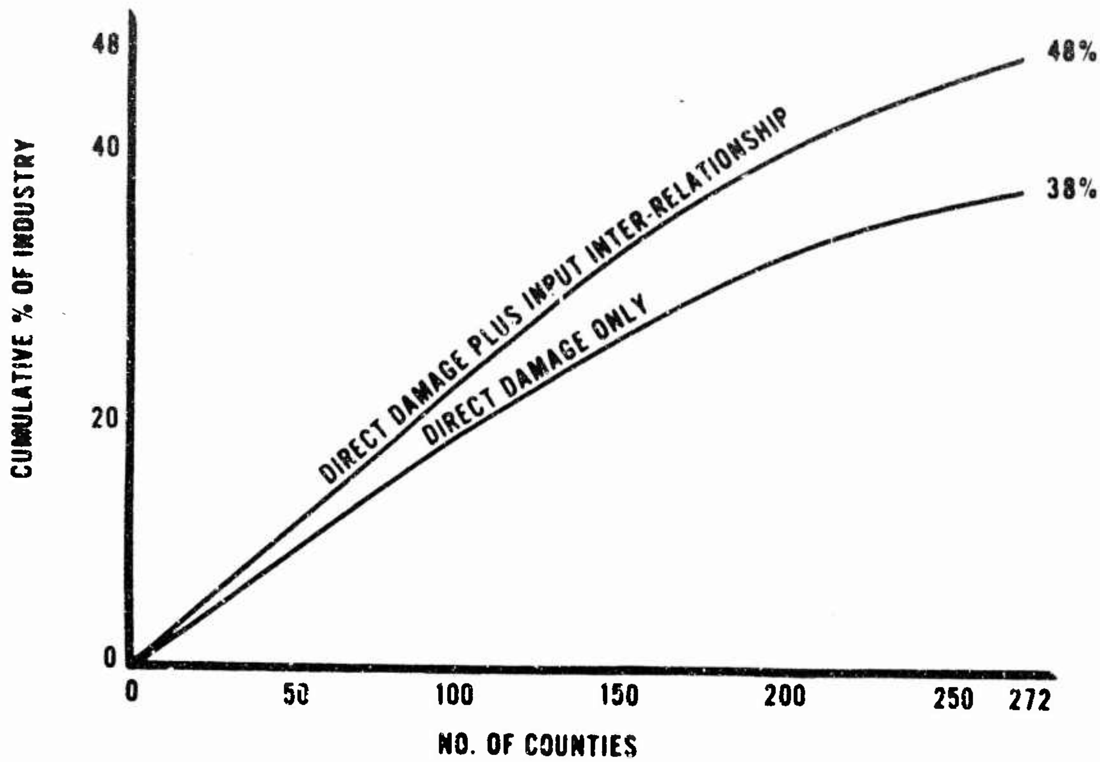


FIGURE 3

SCATTER DIAGRAM OF INDUSTRIAL INTER-DEPENDENCE VS URBAN CONCENTRATION

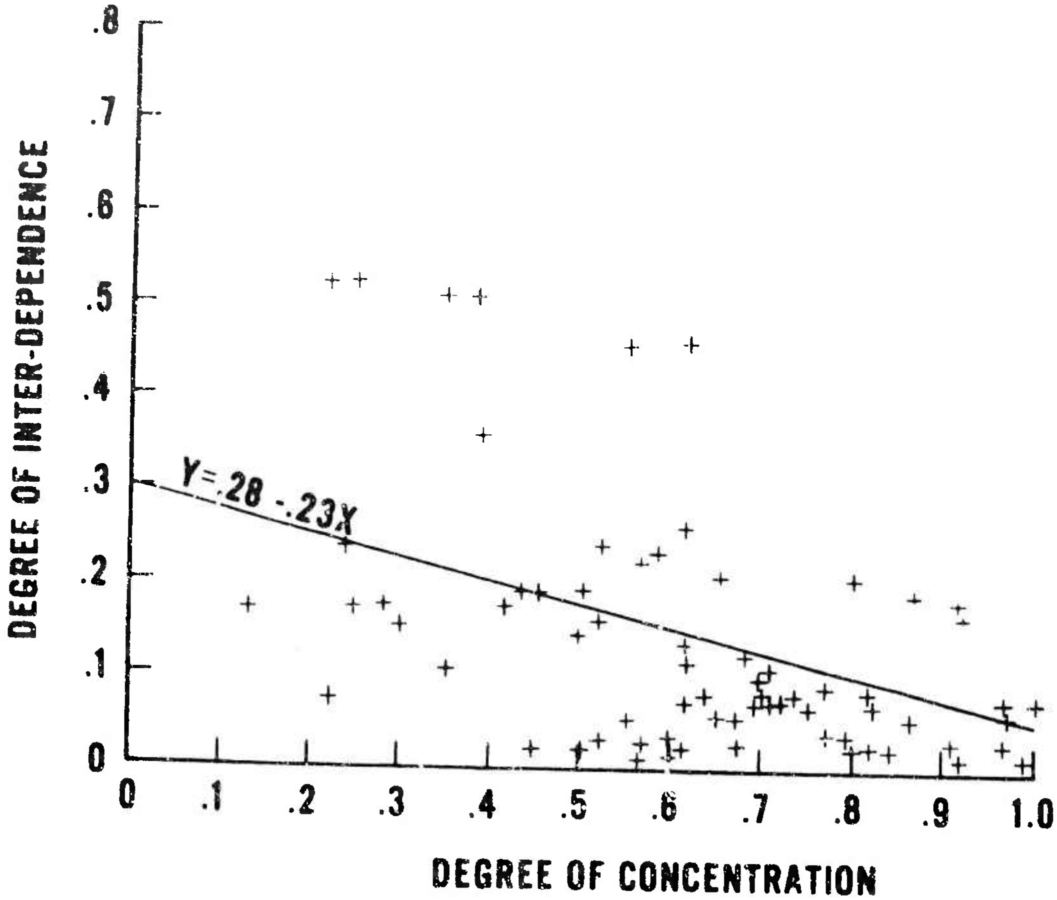


FIGURE 4

NUMBER OF COUNTIES VS

THE CUMULATIVE PERCENTAGES OF POPULATION AND INDUSTRY

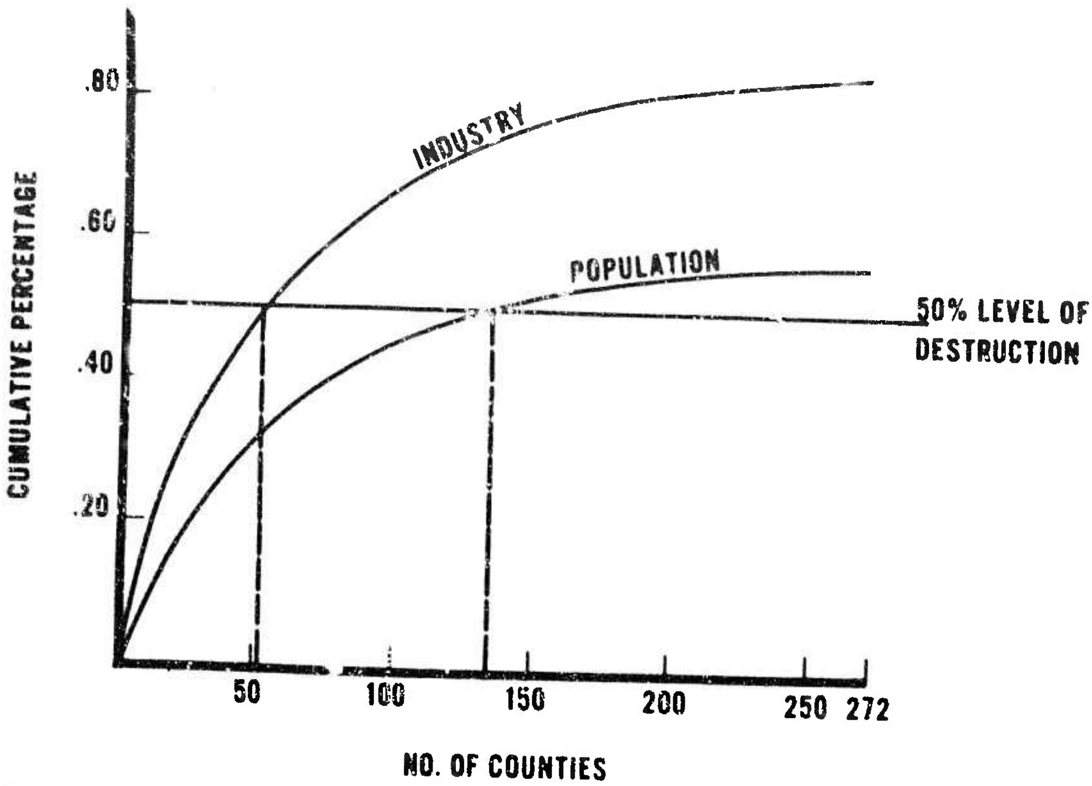


FIGURE 5

ACKNOWLEDGMENT

The author is pleased to acknowledge the following people whose original developmental work at ESSG was the basis for this paper.

Major C. W. Focht, US Army Command and General Staff School, Ft. Leavenworth, Kansas

Mr. Joseph R. Wright, Booz, Allen, and Hamilton Associates, New York, N. Y.

Mr. Ronald C. Schafer, Standard Oil Company of California, San Francisco, California

While not directly reflected in this paper, acknowledgment is made to Mrs. Ann Fishkin and Mr. Miles Letts of the Systems Sciences Corporation, Bethesda, Maryland for their many hours of assistance in adapting the DASH model to our data, and to Mr. Charles Dukes, National Civil Defense Computer Facility, Olney, Maryland for scheduling our use of the CDC 3600 computer.

BIBLIOGRAPHY

1. Business and Defense Services Administration, US Department of Commerce, US Industrial Outlook 1965. US Government Printing Office: Washington, D.C., December 1964.

2. Center for Economic Projection, American Industry in 1976 and 1985 (Projections of Output, Employment, and Productivity). Report No. 64-1. National Planning Association: Washington, D.C., February 1964.

3. Center for Economic Projections, State Population, Net Migration, Labor Force and Industry Employment Trends to 1975. REPS-Rpt No. 65-I. National Planning Association: Washington, D.C., March 1965.

4. Department of Defense, Defense Study Group, Medical Care for Retired Military Personnel and Their Dependents. (A report to the Secretary of Defense.) Washington, D.C., June 1964.

5. Denison, Edward, Sources of Economic Growth in the United States. Committee for Economic Growth: New York, 1962.

6. Department of Defense, Office of Defense Mobilization, Reducing Our Vulnerability to Attack. Washington, D.C., 1957.

7. Dewhurst, J. Frederick and Associates, America's Needs and Resources. New York: The Twentieth Century Fund, 1955.

8. Executive Office of the President, Office of Emergency Planning, PARM System Manual. Vol III, Section 0: Manpower and Occupations. NREC Technical Report No. 17. National Planning Association: Washington, D.C., December 1963.

9. Executive Office of the President, Bureau of the Budget, Office of Statistical Standards, Standard Metropolitan Statistical Areas. US Government Printing Office: Washington, D.C., 1964.

10. Executive Office of the President, Office of Emergency Planning, PARM System. National Resource Evaluation Center, Washington, D.C. Numerous documents with varying dates.

11. Hirsch, W.Z., "Interindustry Relations of a Metropolitan Area," The Review of Economics and Statistics. Vol. XLI, No. 4, Cambridge, Massachusetts, November 1959, pp. 360-369.

12. Office of Business Economics, US Department of Commerce, "The Inter-industry Structure of the United States," Survey of Current Business. US Government Printing Office: Washington, D.C., November 1964.

13. Organization for European Economic Co-operation, Demographic Trends 1956-1976 in Western Europe and in the United States. McGraw-Hill: New York, undated.

14. President of the United States, 1965 Economic Report of the President. US Government Printing Office: Washington, D.C., 1965.

15. Rand Corporation, Economic Viability After Thermonuclear War: The Limits of Feasible Production. RM-3436-PR. Santa Monica, California, September 1963.

16. Rand Corporation, The, Economic Viability After Thermonuclear War. The Limit of Feasible Production. Memorandum RM-3436-PR. Santa Monica, California, September 1963.

17. Rand Corporation, Metropolitan Population to 1985: Trial Projections. Memorandum RM-4125-RC. Santa Monica, California, September 1964.

18. Research Analysis Corporation, The 1959 Soviet Intersectoral Flow Table: Volume I. McLean, Va., November 1964.

19. Stolnitz, George J., Manpower Movements: A Proposed Approach to Measurement. The Johns Hopkins Press: Baltimore, 1964.

20. System Development Corporation, Jovial Training Manual for the 1604B Computer. Falls Church, Va., March 1965.

21. System Sciences, Inc., Preliminary Notes on DASH: A System for Providing a Detailed Assessment of Hazard for Civil Defense Purposes. Bethesda, Md., September 1965.

22. Technical Committee on Industrial Classification, Office of Statistical Standards, Standard Industrial Classification Manual, with SIC Manual Supplement. Washington, D.C., 1957 and 1963, respectively.

23. US Bureau of the Census, Census of Business. Washington, D.C., 1960.

24. US Bureau of the Census, Census of Manufacturers. Washington, D.C., 1960.

25. US Bureau of the Census, Census of Population and Housing--1960. Washington, D.C., 1960.

26. US Bureau of the Census, Census of Tracts--1960. Washington, D.C., 1960.

27. US Bureau of the Census, Current Population Reports. Population Estimates. Series P-25, No. 286. US Government Printing Office: Washington, D.C., July 1964.

28. US Bureau of the Census, US Census of Population: 1960. Occupational Characteristics. Final Report PC(1)-D series. US Government Printing Office: Washington, D.C., 1963.

29. US Intelligence Board, Committee on Documentation, Intelligence Subject Code and Area Classification Code. 2nd edition. Washington, D.C., January 1962.

US Department of Commerce, Office of Business Economics, The 1958 Interindustry Relations Study. Washington, D.C., November 1964.

31. US Department of Commerce, Office of Business Economics, Survey of Current Business. Washington, D.C., November 1964.

32. US Bureau of the Census, Annual Survey of Manufactures 1962. US Government Printing Office: Washington, D.C., 1964.

33. US Department of Labor, Manpower Report of the President and A Report on Manpower Requirements, Resources, Utilization, and Training. US Government Printing Office: Washington, D.C., March 1965.

34. US Bureau of Labor Statistics, Manpower Needs by Industry to 1975. US Government Printing Office: Washington, D.C., 1965.

35. US Bureau of the Census, County Business Patterns, 1959. US Government Printing Office: Washington, D.C., 1960.

36. ________________, County Business Patterns, 1962. US Government Printing Office: Washington, D.C., 1963.

37. ________________, Current Population Reports. Population Estimates. Series P-25, No. 298. US Government Printing Office: Washington, D.C., 12 February 1965.

38. ________________, Current Population Reports. Technical Studies. Series P-23, No. 10. US Government Printing Office: Washington, D.C., 5 December 1963.

39. Rand Corporation, Metropolitan Populations to 1985: Trial Projections. Memorandum RM-4125-RC. Santa Monica, California, September 1964.

40. US Bureau of Labor Statistics, Special Labor Force Report. Labor Force Projections for 1970-80. Report No. 49. US Government Printing Office: Washington, D.C., 1965.

41. Center for Economic Projection, State Population, Net Migration, Labor Force and Industry Employment Trends to 1975. Report No. 65-1. National Planning Association: Washington, D.C., March 1965.

42. Rand Corporation, US Census of Population: 1960. Characteristics of the Population. Number of Inhabitants. Vol. 1, Part A. US Government Printing Office: Washington, D.C., 1961.

43. US Bureau of the Census, US Census of Population: 1960. Selected Area Reports. Standard Metropolitan Statistical Areas. Final Report PC(3)-1D. US Government Printing Office: Washington, D.C., 1963.

44. ________________, US Census of the Population: 1960. Subject Reports. Occupation by Industry. Final Report PC(2)-7C. US Government Printing Office: Washington, D.C., 1963.

45. ________________, US Censuses of Population and Housing: 1960. Census Tracts. Final Report PHC(1)-147. US Government Printing Office: Washington, D.C., 1962.

46. ________________, US Census of Population: 1950. Special Reports. Industrial Characteristics. Vol. IV, Part 1, Ch D. US Government Printing Office: Washington, D.C., 1955.

47. ________________, US Census of Population: 1950. Special Reports. Occupational Characteristics. Vol. IV, Part 1, Ch B. US Government Printing Office: Washington, D.C., 1956.

48. ________________, US Census of Population: 1950. Special Reports. Occupation by Industry. Vol. IV, Part 1, Ch C. US Government Printing Office: Washington, D.C., 1954.

49. ________________, US Census of Population: 1960. Characteristics of the Population. Number of Inhabitants. Vol. I, Part A. US Government Printing Office: Washington, D.C., 1961.

50. US Bureau of the Census, US Census of Population: 1960. General Social and Economic Characteristics. United States Summary. Final Report PC(1)-1C. US Government Printing Office: Washington, D.C., 1962.

COST-EFFECTIVENESS TECHNIQUES FOR A THEATER ARMY COMMUNICATIONS SYSTEM

Dr. Richard L. Crawford
Combined Arms Research Office
Booz, Allen Applied Research Inc.
Fort Leavenworth, Kansas
Cognizant Agency: U.S. Army Combat Developments Command

Introduction

The design of a Theater Army Communications System is to be based upon minimum essential communications requirements. The requirements are to be satisfied using a minimum cost communications system. This paper describes the cost-effectiveness techniques that were developed for the design of the Theater Army Communications System (TACS-70). 1/

A minimum cost communications system can be designed based upon the channel requirements developed from the user requirements. However, it is important that various significant factors which serve to define the system's cost and performance be properly considered in designing a "minimum cost" communications system. In addition, it should be emphasized that a cost-effectiveness analysis is not a decision process, but is an aid to the communications system designer. The analysis seeks to quantify what can be logically calculated so that the decision maker knows the extent to which intuitive judgment must be used in making a decision. In the areas of intuitive judgment, the decision maker must assign weights to such effectiveness factors as survivability, electromagnetic compatibility, electronic countermeasures sensitivity and excess channel requirements. All these factors are influenced by the threat situation. Thus, a minimum cost communications system implies that cost-effectiveness techniques have been applied and that the system is the least expensive for the desired degree of effectiveness.

The cost-effectiveness relationships developed for the TACS-70 system dealt directly with the following cost-effectiveness factors: dollar cost of the equipment (initial and operational), channel capacity of the links, range coverage, survivability and switching. The electromagnetic compatability and electronic countermeasures sensitivity of the various equipment were evaluated separately and are not included as effectiveness measures in this paper. These factors influence the choice of equipment for a particular link and would show up in the costs of the links.

Since a large number of parameters affect both cost and effectiveness, and all possible permutations of the parameters would give a very large number of theoretically possible solutions, it was necessary to follow the following ground rules:

- Only equipment and cost data included in the TACS-70 equipment list were considered.
- The quality of the system was maintained.
- Equipment that is normally organic to a theater unit, other than signal, was not considered.
- Mobility and installation time were not considered since all the equipment is truck-mounted and has comparable mobility.
- All transmission ranges used were the average given in the equipment list.
- The impact on effectiveness of fade-outs, noise bursts, and temporary loss of signal was not considered.
- On all equipment, the maximum channel capacity was used.

Cost-Effectiveness of the Links

In the TACOM [2] (tactical communications study) list of equipment, there are 21 different combinations of basic equipment to form terminals or relays. This list did not necessarily apply to the TACS-70 study and new configurations were assembled to reduce costs or to satisfy new operational requirements.

The first step in the analysis was to define the equipment that would be used. Curves similar to the one shown in Figure 1 on multiplex equipment were developed for radio-relay terminals, radio relays, wire terminals and tropospheric scatter terminals. These curves included Standard-A and developmental equipment that would be available in the time frame of the study. Gaps in the channel coverage were filled by configuring new terminals or relays from the basic radios and multiplex equipment.

A communications network is described in terms of network nodes (major network transmission centers) and links (the communication means between nodes). A cost-effectiveness analysis of the links determines the least costly way to provide a given channel capacity between nodes.

The effects of the distance parameter upon the three competitive

means of communication are shown in Figure 2. Figure 2 shows the cost-distance curves for twelve channels. Similar curves were drawn for 24, 48 and 96 channel capacity equipment. The cost of zero miles is the cost of two terminals. The stepwise jumps for Radrel (radio-relay) are the costs for relay stations (procurement cost only). The gradual rise for Wire is due to the cost of the wire and pulse restorers. The jump every forty miles is the cost of the AN/TCC-64, the PCM pulse repeater station. The jumps in the Tropo (tropospheric-scatter) curves occur at the maximum distance the Tropo terminals can be separated and represent the cost of two more terminals to be used as relays.

The four sets of cost-distance curves were then used to develop the cost-effectiveness curves as shown in Figure 3. For each link distance, in ten mile steps, there is a separate curve. These curves show the minimum cost for each channel capacity, in twelve channel jumps, for the three competitive systems. As a tool for the network designers, these curves show the cost differentials between different channel capacities. For example, at a fifty-mile separation, it would cost only $4,000 more to go from 24 channels to 48 channels, while it would cost $150,000 to step up to 60 channels from 48, at an eighty-mile separation.

Figure 4 converts the cost-effectiveness curves into bar graphs to indicate which of the three communication means is the least costly for a given node separation and channel capacity. The slanting lines indicate a region of approximately equal cost between the two means on either side of the line. The cross-hatched areas indicate that the next greater channel capacity is less costly at these ranges than any of the three communication means at this channel capacity.

The results so far have only considered acquisition costs. Figure 5 shows how the cost increase per year per relay station that is added to the link. These costs are the operational and maintenance (O&M) costs of the relay equipment and personnel. If the system designer wants to include the O&M costs, he decides on the time period for assigning these costs and includes them in the cost-distance curves. The O&M costs of the terminals do not affect the slope of the curves. However, they do affect the crossover points (the point of equal cost) of the competing systems and must be included. The cost-effectiveness curves and the bar graphs that are drawn from the cost-distance curves that include O&M costs show that, with increased time, the Tropo system becomes more desirable, especially with link distances greater than forty miles.

Cost-Effectiveness of the System

The next step in the design of a communications system was to construct a complete network based upon the cost-effectiveness curves that were developed for the links. The network must satisfy the user

communications requirements (UCR) of the Theater Army. To satisfy the UCR, a common model was defined. The model proposed was based upon the theater operations associated with twelve division three corps force configurations. To further define the model, the number of nodes must be specified. The number used for concept comparison was thirty. The placement of the n s is flexible and, to satisfy the needs of the users, they can be placed anywhere within the boundaries of the Theater at the discretion of the network designer.

The optimal network configuration is strongly dependent upon the flow of traffic. This dependence will show up in the survivability index, the redundance required, the number of channels per link and the average time required to complete a call.

A wide variety of networks can be constructed between a set of given nodes. However, they all factor into two components: centralized (or star) and distributed (or grid or mesh). The centralized network is obviously vulnerable, since destruction of a single central node destroys communication between the end stations. In practice, a mixture of star and mesh components is used to form communication networks.

Six different distributed network concepts were proposed using the thirty nodes in the model. The concepts are shown in Figure 6 to 11. The six networks differ in redundancy and topology. The variations were chosen to compare the effects of redundancy and topology on survivability and switching effectiveness. Redundancy of links is defined as the number of links used minus the minimum number necessary to join every node in a network. Thus, redundancy is given by:

$$R = b - (n-1) \tag{1}$$

where b is the number of links used and n is the number of nodes in the network. The redundancy of each concept is shown in the figures.

Concept 1, Figure 6, is the area system proposed by the TAACS study. 3/ Concept 2, Figure 7, is an area system with a redundancy of 35. Concept 3, Figure 8, has the same node location as Concept 2, but with a smaller redundancy and different link connections. Concept 4, Figure 9, is similar to Concept 1, but with a smaller redundancy. Concept 5, Figure 10, is an adaptive concept. The adaptive principle uses two links in the central path to connect each node. If one of the central nodes is destroyed, then the two adjacent central nodes could establish new links with the nodes opposite the destroyed node in the outer paths. The new links are created at the discretion of the signal control officer who is cognizant of the network status after a raid. The reason for dual links in the central path was that

the range of Tropo systems is dependent upon channel capacity. The smaller capacity system has a greater range and, hence, could establish a diagonal link. The other concepts could also use the adaptive principle. However, only Concept 5 used the adaptive principle to illustrate the advantages of responsive signal control action. Concept 6, Figure 11, was designed to serve large population centers in a typical theater area with long direct links.

A criterion of survivability is necessary to a consideration of the synthesis of a communication network which allows communication after partial destruction. This criterion, or index of survivability, can be used to determine which of the concepts proposed has the maximum survivability in case one or more of the nodes or links are destroyed. Baran 4/ suggests the ratio:

$$I_1 = \frac{\text{Largest number of nodes in one group after destruction}}{\text{Number of nodes connected together before destruction}} \quad (2)$$

as the index of survivability. Deo 5/ proposes the following ratio:

$$I_2 = \frac{\text{Number of pairs of nodes remaining connected after destruction}}{\text{Number of pairs of nodes in connection before destruction}} \quad (3)$$

The first index, I_1, is a measure of the ability of the surviving nodes to operate together as a coherent entity after destruction. This means that small groups of nodes isolated from the single largest group are considered to be ineffective. Deo's measure acknowledges meaningful communication between the nodes in the small groups after destruction.

Neither of these indexes are adequate. A true measure of survivability should take into account the traffic flow in a network. Just because the i^{th} node is connected to the j^{th} does not mean that a relevant message between the two can be sent. The connection may have been established as forward flow route. A relevant message is one addressed to the j^{th} node for action and not one addressed to the j^{th} node for further flow by an alternate means. The index of survivability used is the ratio:

$$I = \frac{\text{Number of messages reaching their destination after destruction}}{\text{Number of messages reaching their destination before destruction}} \quad (4)$$

Figure 12 shows the survivability curves of the six concepts.

The curves were obtained by a Monte-Carlo simulation of node destruction. The traffic flow for this simulation was generated from what is expected in the theater. The survivability index, eq. (4), was calculated from node destruction in ten percent steps for each concept. For each step, the Monte-Carlo simulation randomly determined the nodes destroyed. The survivability index was taken as the average of ten trials. The curves shown in Figure 12 are the best fit to the plotted data. To quantify survivability, a figure called the percent of best possible survivability was obtained for each curve. The best possible line is for a network with maximum redundancy. A network with maximum redundancy has every node connected to every other node. The curves are integrated and the integrals are then compared with the integral of the best possible survivability line in an arbitrary assessment of "expected" survivability. The critical point is between thirty to forty percent node destruction. With the exception of Concept 5, the higher the redundancy, the greater the survivability. Concepts 2, 3 and 5, in addition to having greater redundancy, benefit due to their shapes. The nine by three matrix shape of Concepts 1 and 4 is easier to cut than the six by five matrix shape of Concepts 2, 3 and 5.

The technique for determining survivability can be adapted to computer analysis using the traffic flow that is generated from the user communication requirements data.

The Theater Army has perhaps 150,000 to 200,000 men deployed throughout an area of 55,000 square miles. By necessity, a highly complex communications network is required, covering the entire area and serving all users, in which a user can be switched to any other user and connected via the communications channels constituting the network. Therefore, a large number of switching requirements exist in the Theater Army. This switching, at the present time, is accomplished by manually operated switchboards. The Army's experience with manual operation parallels that of the commercial telephone companies; i.e., operators are expensive, have limited efficiency and often cannot provide the speed of service required. It is assumed that an operator can handle a local call in about fifteen seconds. Long-distance calls, however, present a serious problem, because the originating operator must stay with a call through three to five intermediate switchboards until the desired subscriber is connected. The operator holding time then is increased to about 75 seconds per call. Therefore, any network concept that can decrease the number of intermediate switches in a long-distance call will increase the effectiveness of the communications system.

Switching effectiveness of the network concepts can be evaluated from a smooth curve drawn through a histogram of the number of links

required to complete the total daily number of long-distance call in the theater. Figure 13 shows the curves of only one node, the center Field Army access point, with an idealized traffic flow of one message to each of the other 29 nodes. This assumption was based on a parallel computer simulation study 6/ which showed that the total theater traffic tends to be uniformly distributed over the theater area. The histogram for the complete network can be obtained from the computer program that was developed. In order to determine the distribution of traffic, the program only complies that traffic which has passed over N or fewer links. Thus, by performing program runs with increasing values of N and subtracting the previous N-1 run, the amount of traffic that has passed over N links can be obtained. As shown in Figure 13, the mean number of links to reach all nodes is 3.4 for Concepts 2, 3 and 5, 5.33 for Concepts 1 and 4, 2.85 for Concept 6. An average of two more intermediate switches is then required for a long-distance call in Concepts 1 and 4.

There are approximately 12,000 long-distance calls per hour in the busy hours. If there was an average reduction of two intermediate switches, thirty seconds per call could be saved for a total savings of 360,000 seconds or 6,000 call minutes. This savings in call minutes would mean that the total number of operators throughout the network could be reduced and the average speed of service would be faster by thirty seconds. Decreasing the intermediate switching also reduced operator release time. Under heavy operator load conditions, the release time is approximately fifteen seconds. If two intermediate switches are saved, there is a total savings of one minute per long-distance call. Thus, during the busy hour, there is a savings of 12,000 call minutes in Concepts 2, 3 and 5 over Concepts 1 and 4. This savings in call-minutes will mean the trunk groups in Concepts 1 and 4 must be engineered for considerably greater capacity than is required to handle the actual traffic. Speed of service is the measure of effectiveness and the extra capacity and operators required will show up in the cost of the system.

The simulation study 6/ computer program develops the minimum number of channels required for each link from the UCR data for each concept. Then after all design parameters are considered for a network, the designer refers to the cost-effectiveness curves for links and selects the type and channel capacity for each link. The links are defined by type, distance, channel capacity, relays required and cost. The total cost and manpower required are summed for each concept.

Figure 14 shows the cost-effectiveness curve of the communications system with survivability used as a measure of effectiveness. The curve is drawn through the points of decision. Concepts 6 and

2 are above the curve. They are ruled out because there is a concept that is less expensive and more effective. The costs calculated for these concepts were based upon a preliminary assignment of channel capacity to the links and node location.

Figure 15 shows the cost-effectiveness curve with operational and maintenance costs for one year. In this case, the decision curve does not include Concept 3, because the increased O&M costs make it more expensive and less effective than Concept 5. Concept 3 has more relays and consequently more men than Concept 5.

Conclusions

The cost-effectiveness techniques that were presented emphasize the importance of the system designer assigning relative weights to the parameters that affect cost and effectiveness. The designer must make a decision on the following trade-off factors:

- The degree of survivability-redundancy.
- Switching effectiveness.
- The amount of time to assign O&M costs.
- The grade of service by excess channel allocations.

Acknowledgment

The author would like to express his appreciation to Messrs. L. Northrop and C. Raphun of the Communications-Electronics Agency, Fort Huachuca, Arizona for their evaluation of the cost-effectiveness techniques and their helpful suggestions.

References

1. Combat Developments Study Directive: "Theater Army Communication System Requirements, 1965-1970 (TACS-70) (U)", U.S. Army Combat Developments Command, Fort Belvoir, Virginia.

2. "Field Army Requirements for Tactical Communications (TACOM) (U)", November 1964, Final Draft, U.S. Army Combat Developments Command, Fort Belvoir, Virginia.

3. "Theater Army Area Communication System." Office of the Chief of Communication-Electronics Headquarters, Department of the Army, April 1964.

4. P. Baran, "On Distributed Communications: I. Introduction to Distributed Communications Networks," Rand Memorandum RM-3420-PR, August 1964.

5. N. Deo, "On Survivability of Communications Systems," IEEE Transactions on Communication Technology, pp. 227-228, Com-12, December 1964.

6. T. N. Truske, "The Simulation and Evaluation of Traffic Loads and Communications Networks in a Theater Area," 1966 U.S. Army Operations Research Symposium, March 1966.

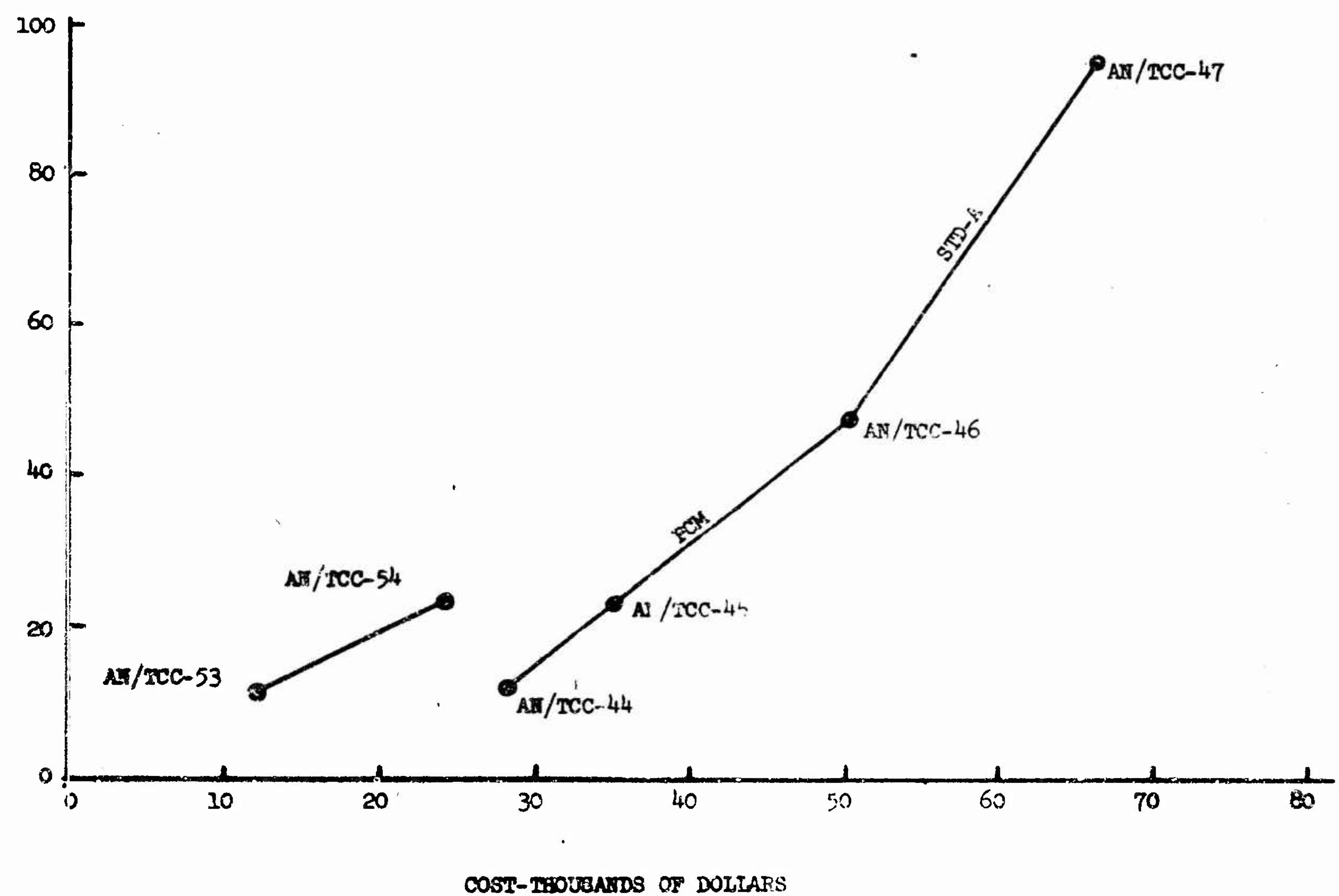


FIGURE 1 MULTIPLEX EQUIPMENT

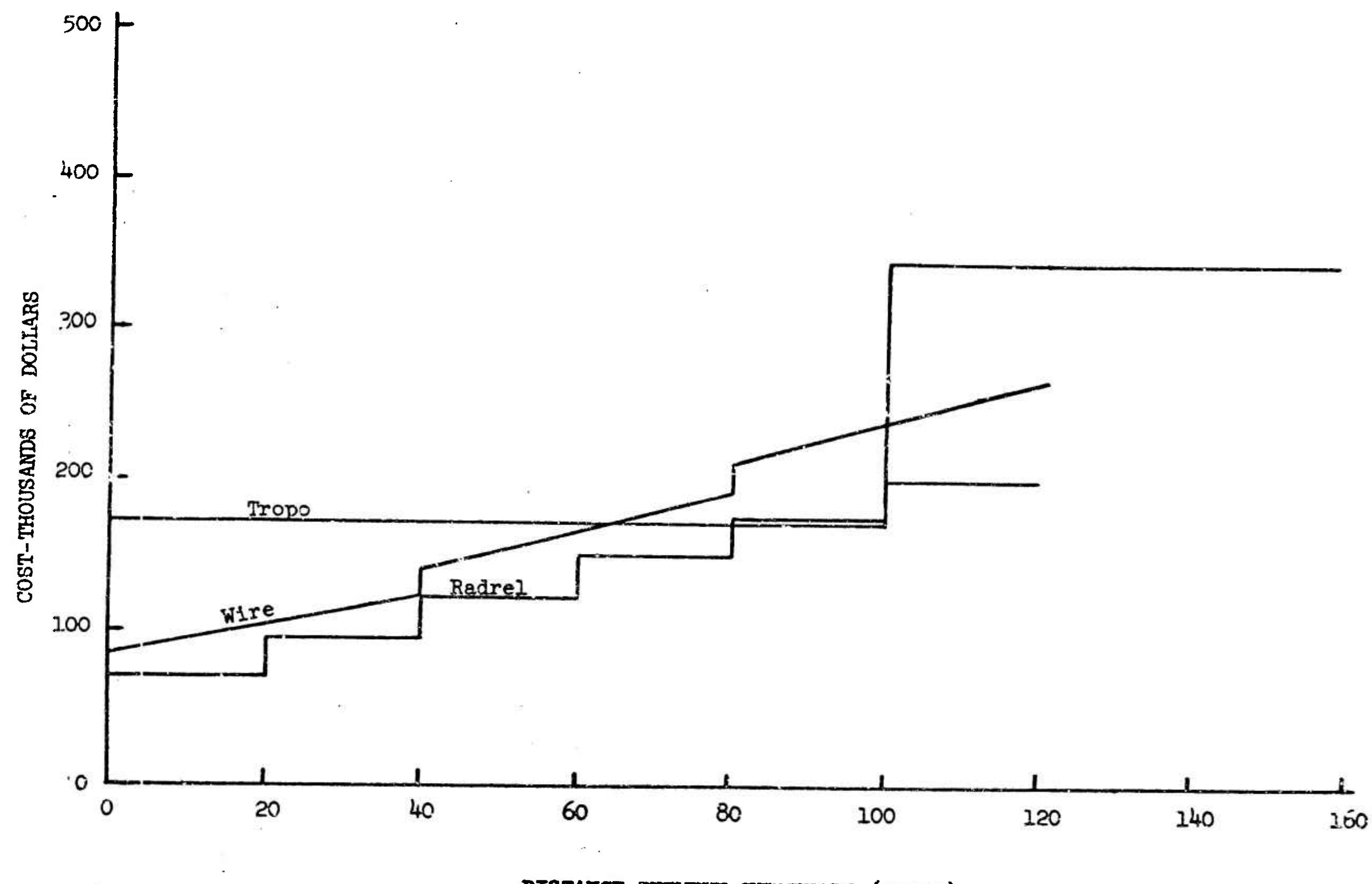


FIGURE 2 COST COMPARISON WIRE VS RADREL VS TROPO 12 CHANNELS

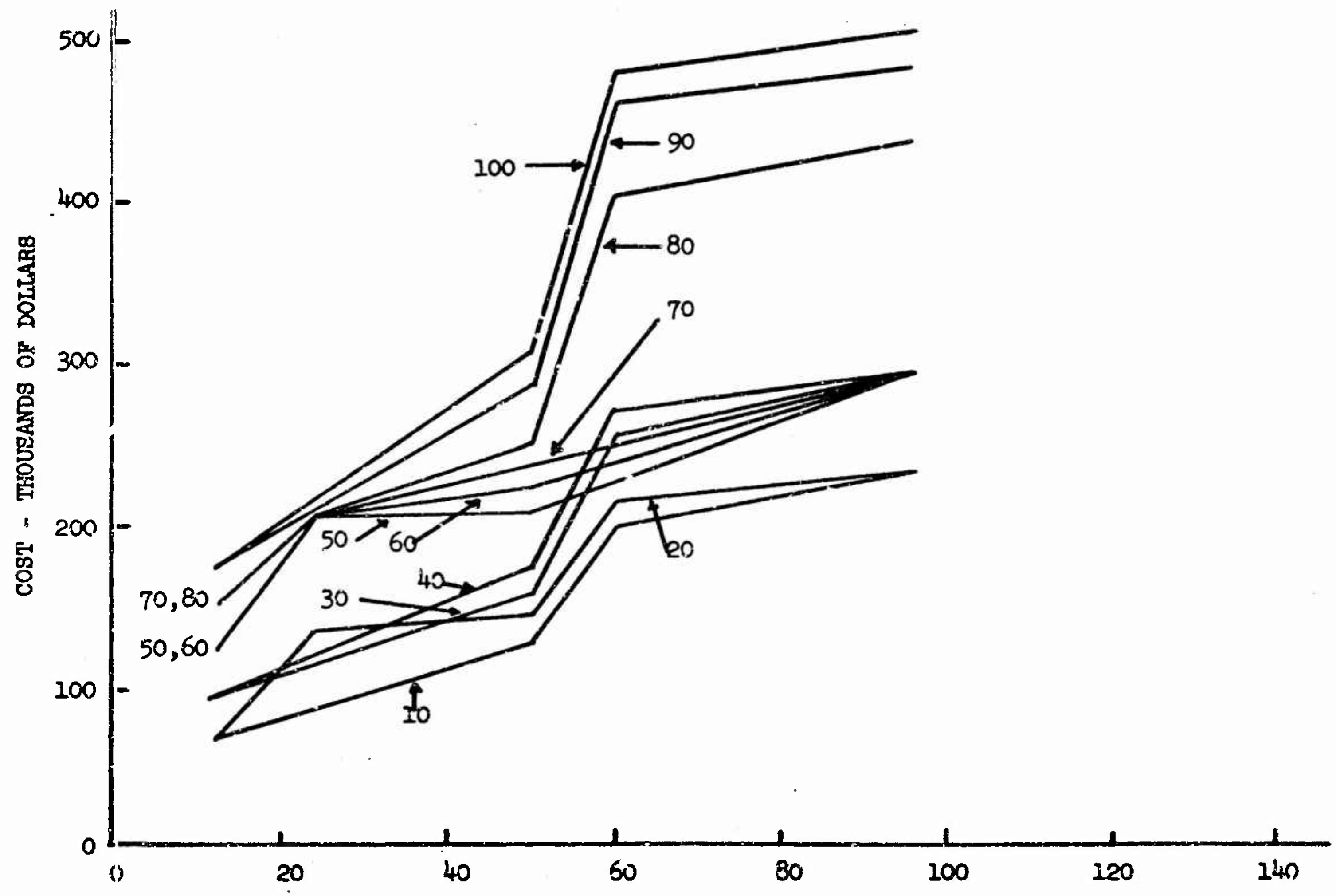


FIGURE 3 COST-EFFECTIVENESS CURVES FOR LINK DISTANCES

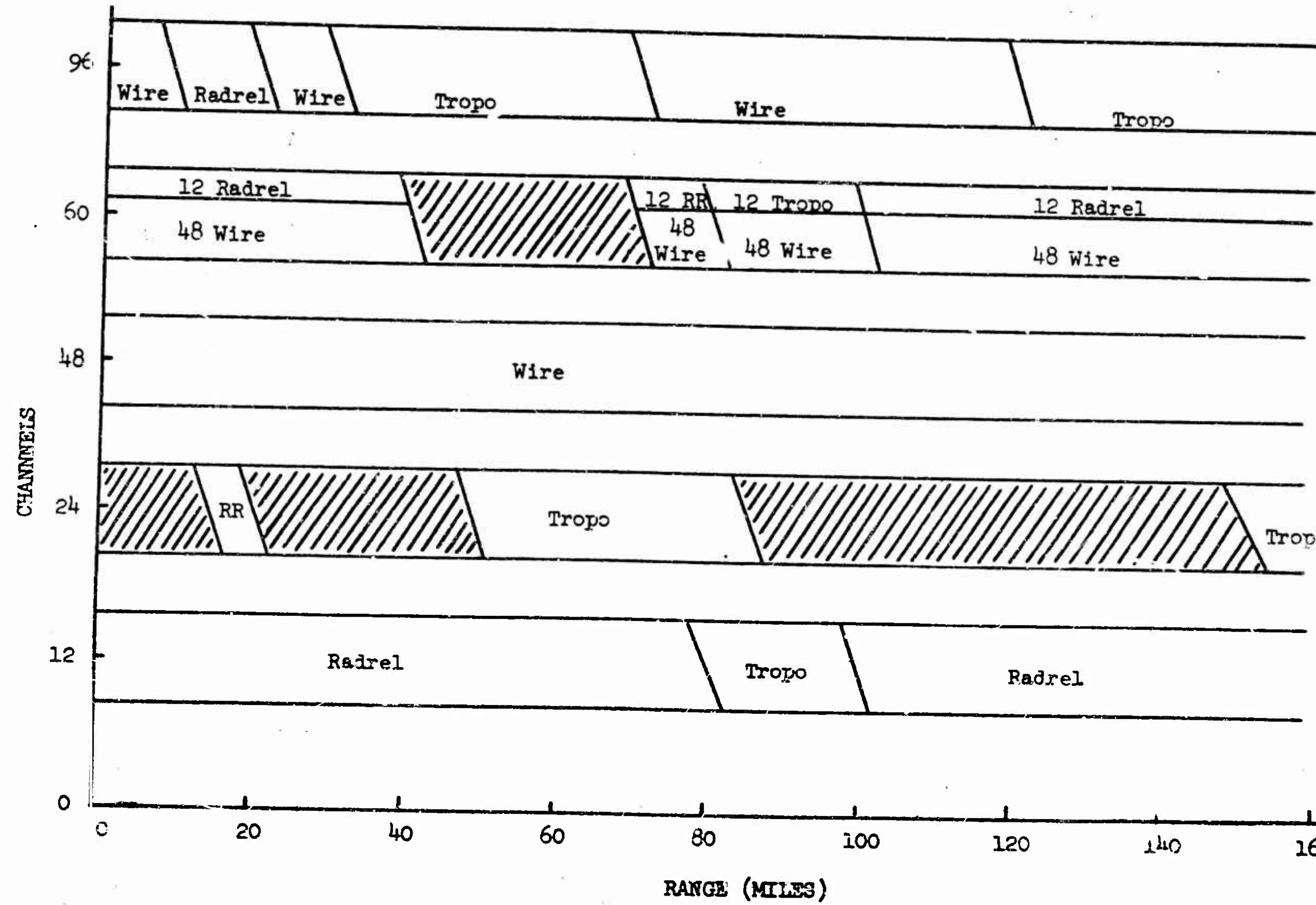


FIGURE 4 EQUIPMENT DESIGNATION FROM COST - EFFECTIVENESS CURVES

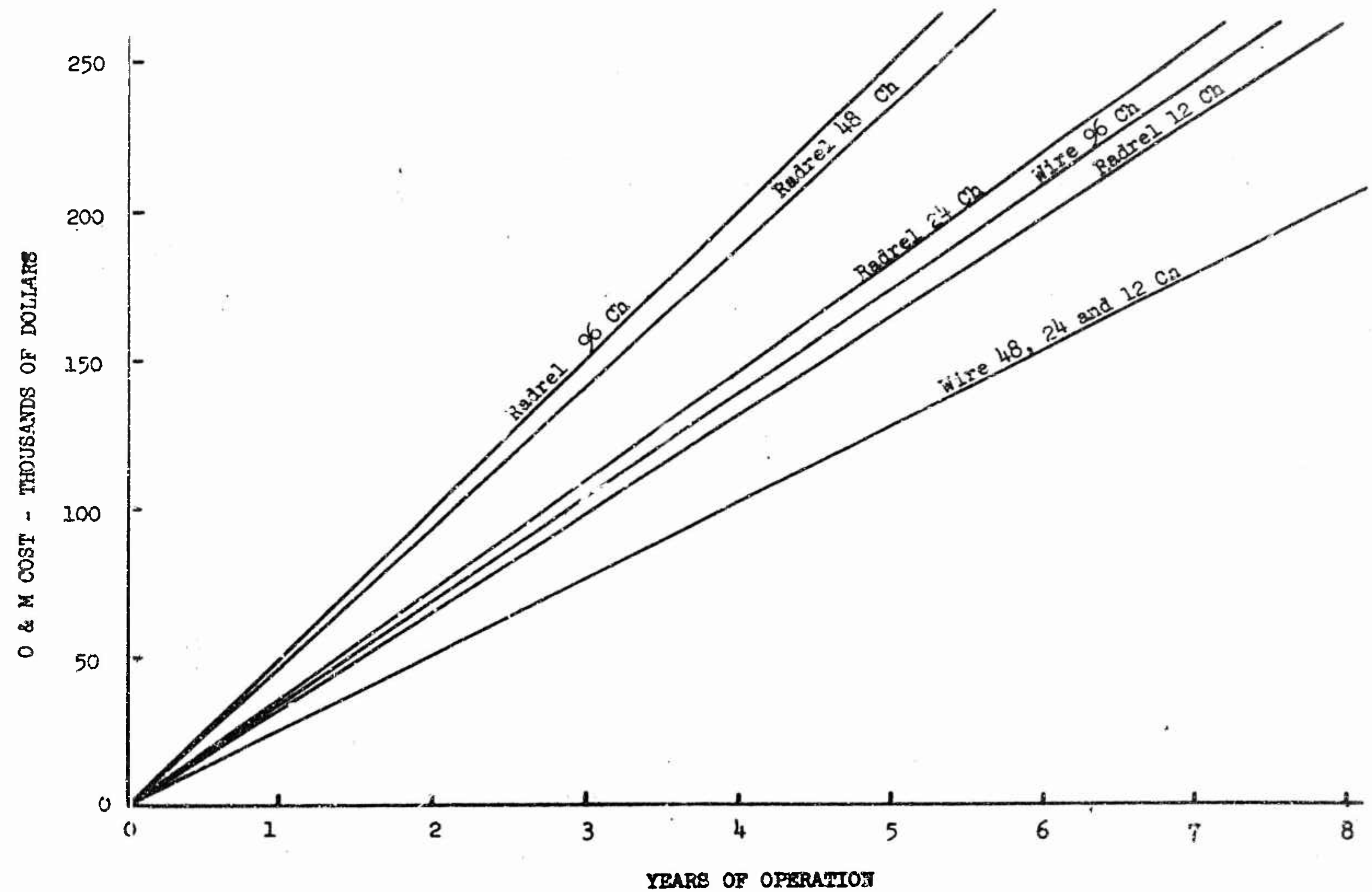


FIGURE 5 OPERATIONAL AND MAINTENANCE COSTS OF RELAY STATIONS

$n = 30$
$b = 46$
$R = 17$

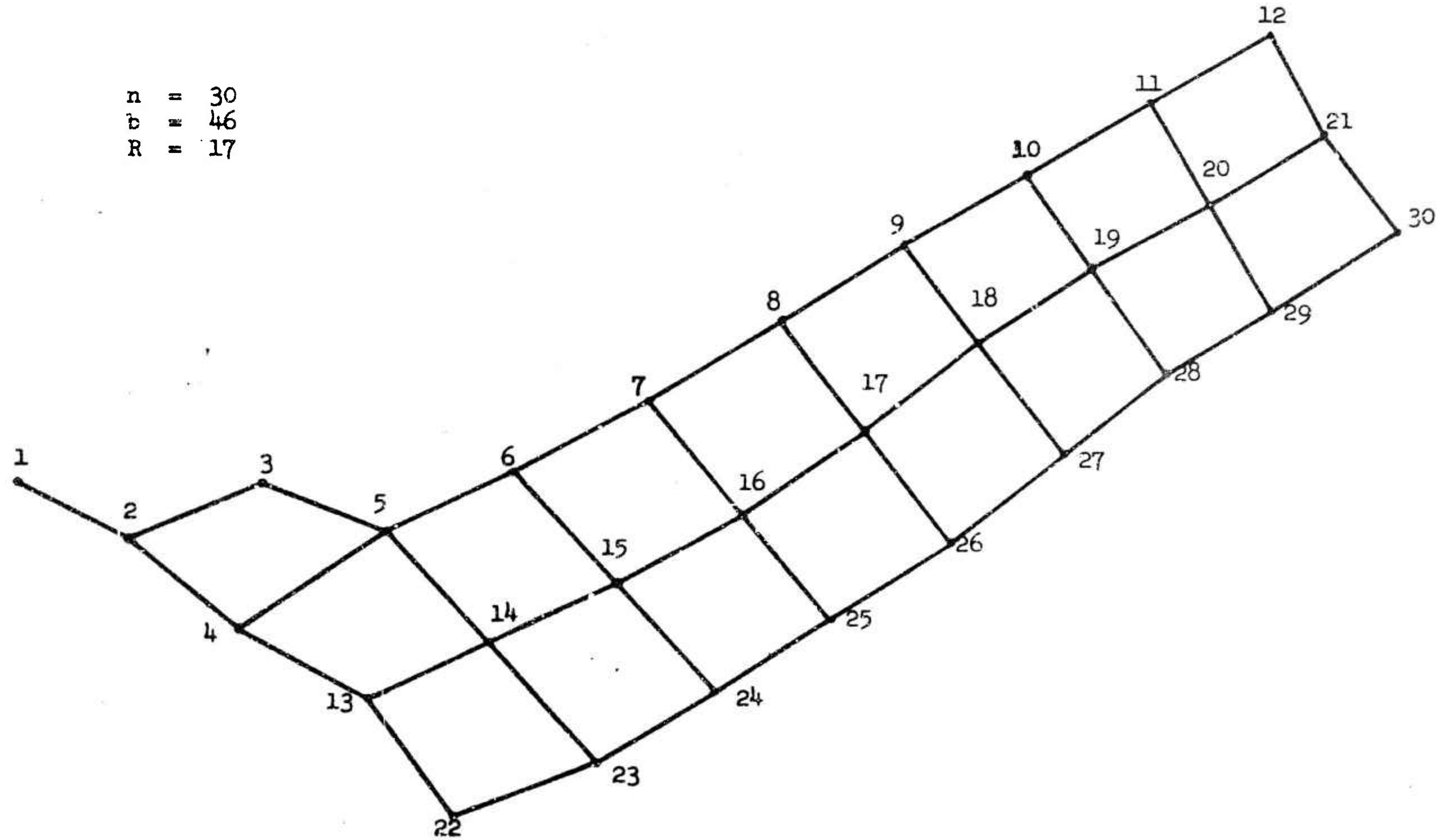


Figure 6 Concept 1

n = 30
b = 64
R = 35

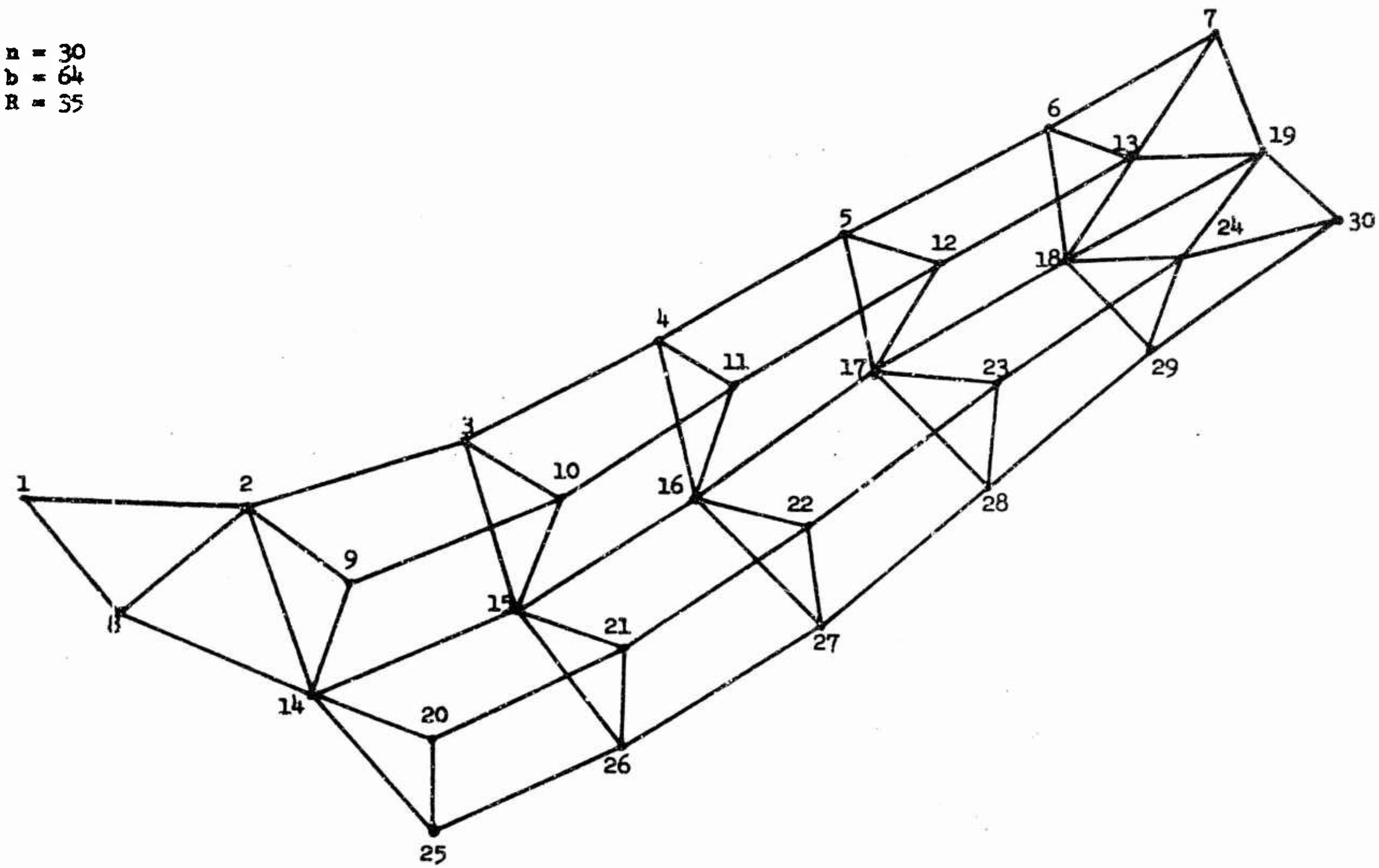


Fig. 7 Concept 2

n = 30
b = 55
R = 26

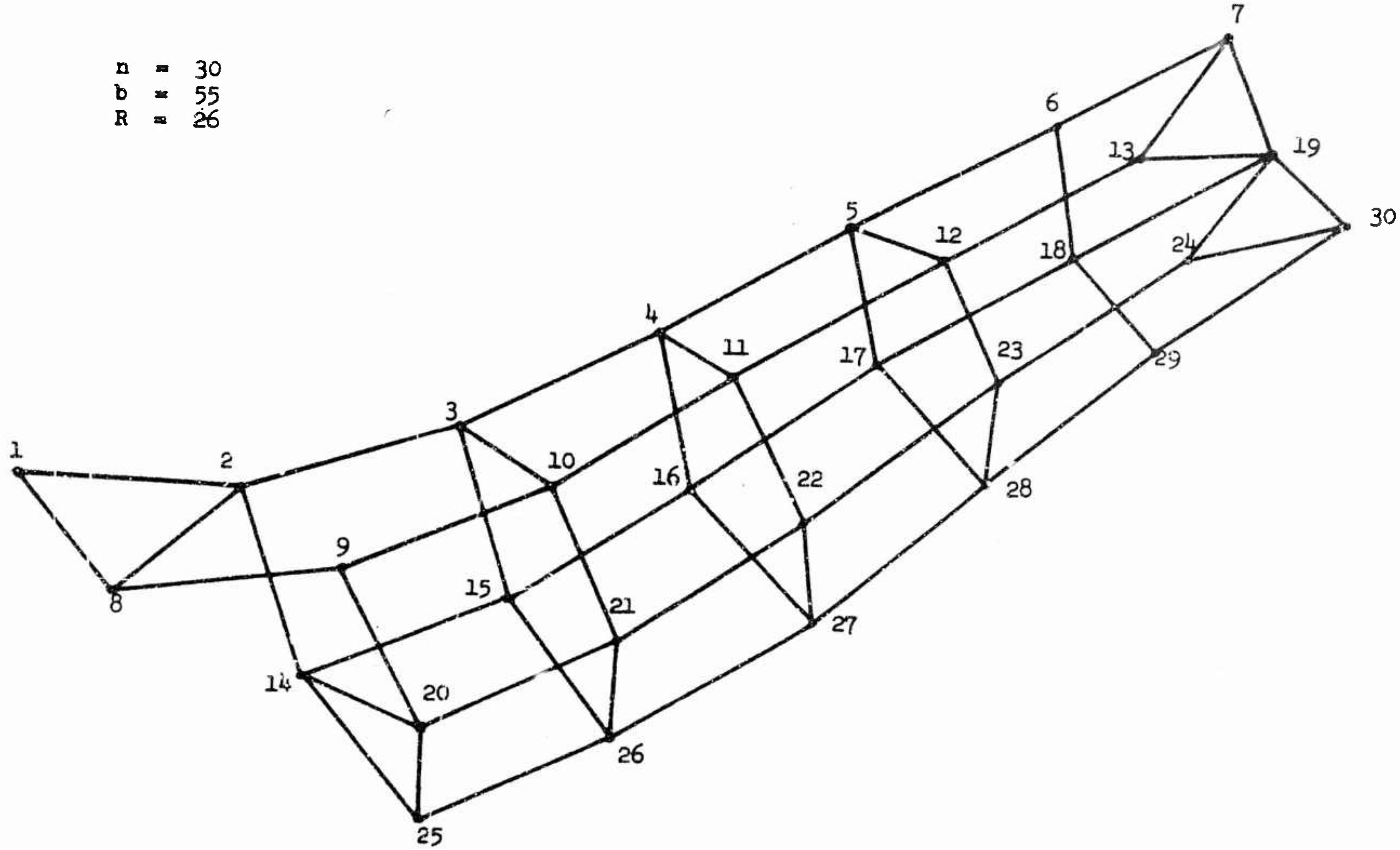


Fig. 8 Concept 3

$n = 30$
$b = 37$
$R = 8$

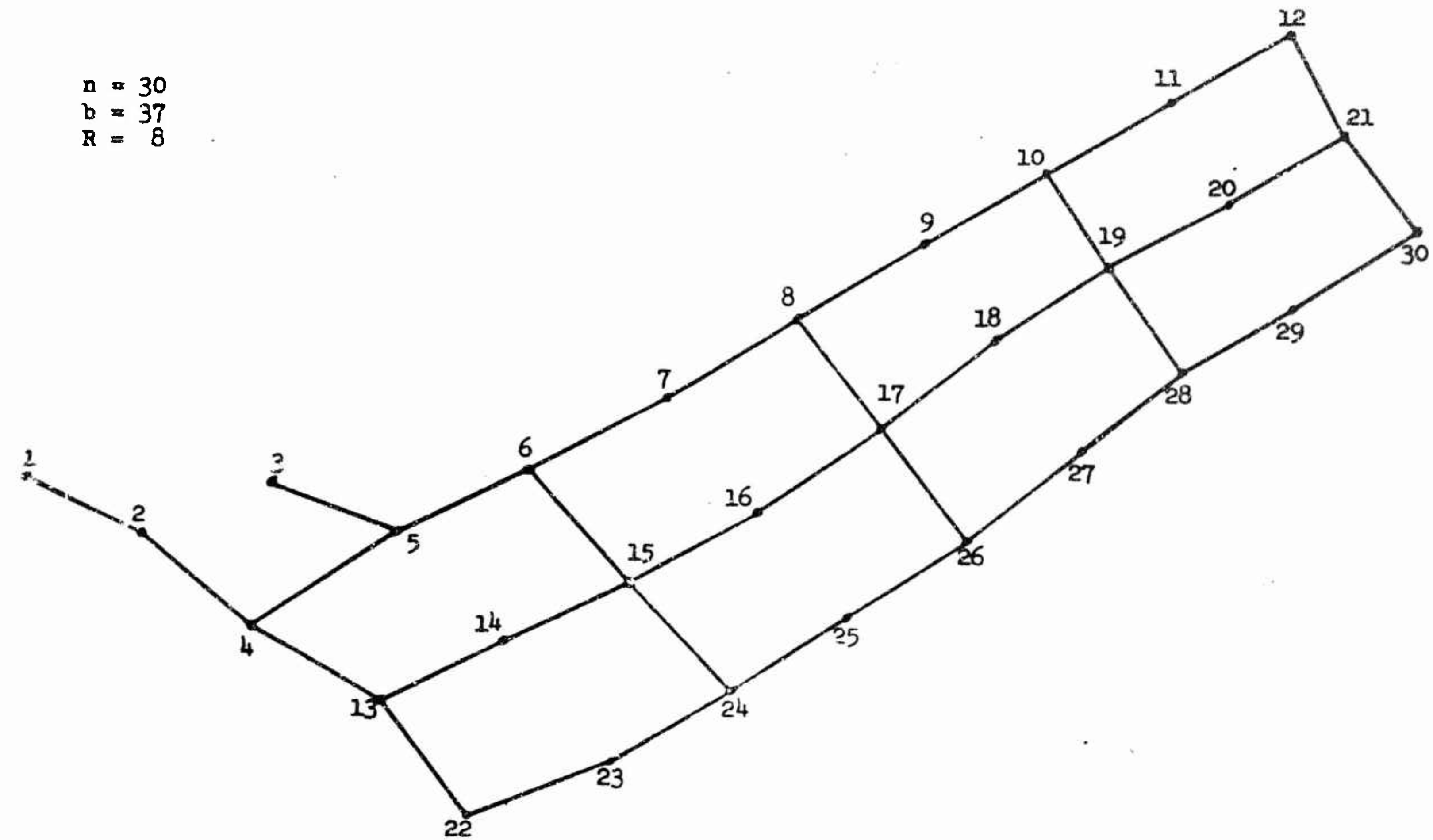


Fig. 9 Concept 4

$n = 30$
$b = 54$
$R = 25$

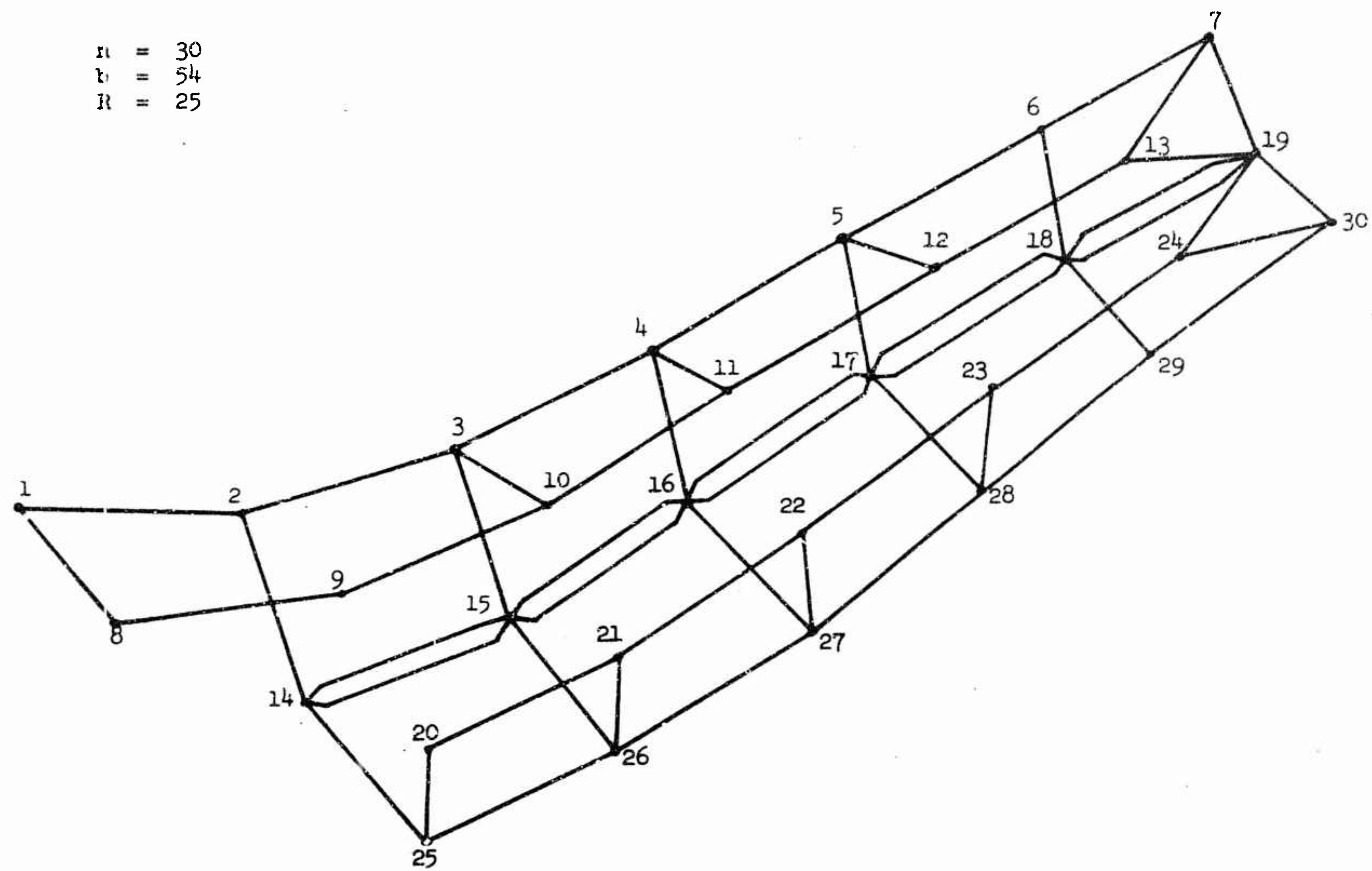


Fig. 10 Concept 5

$n = 28$
$b = 54$
$R = 27$

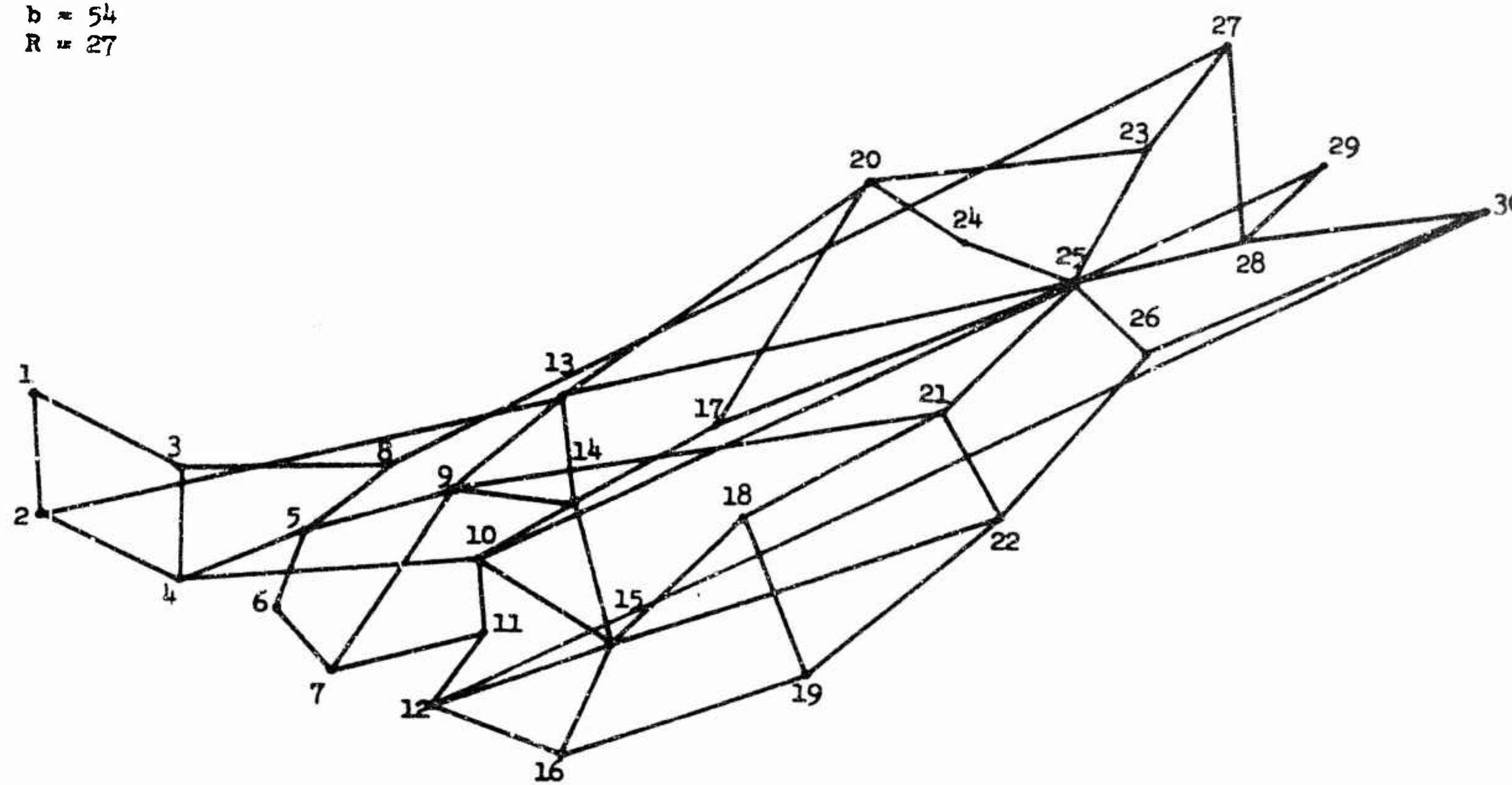


Fig. 11 Concept 6

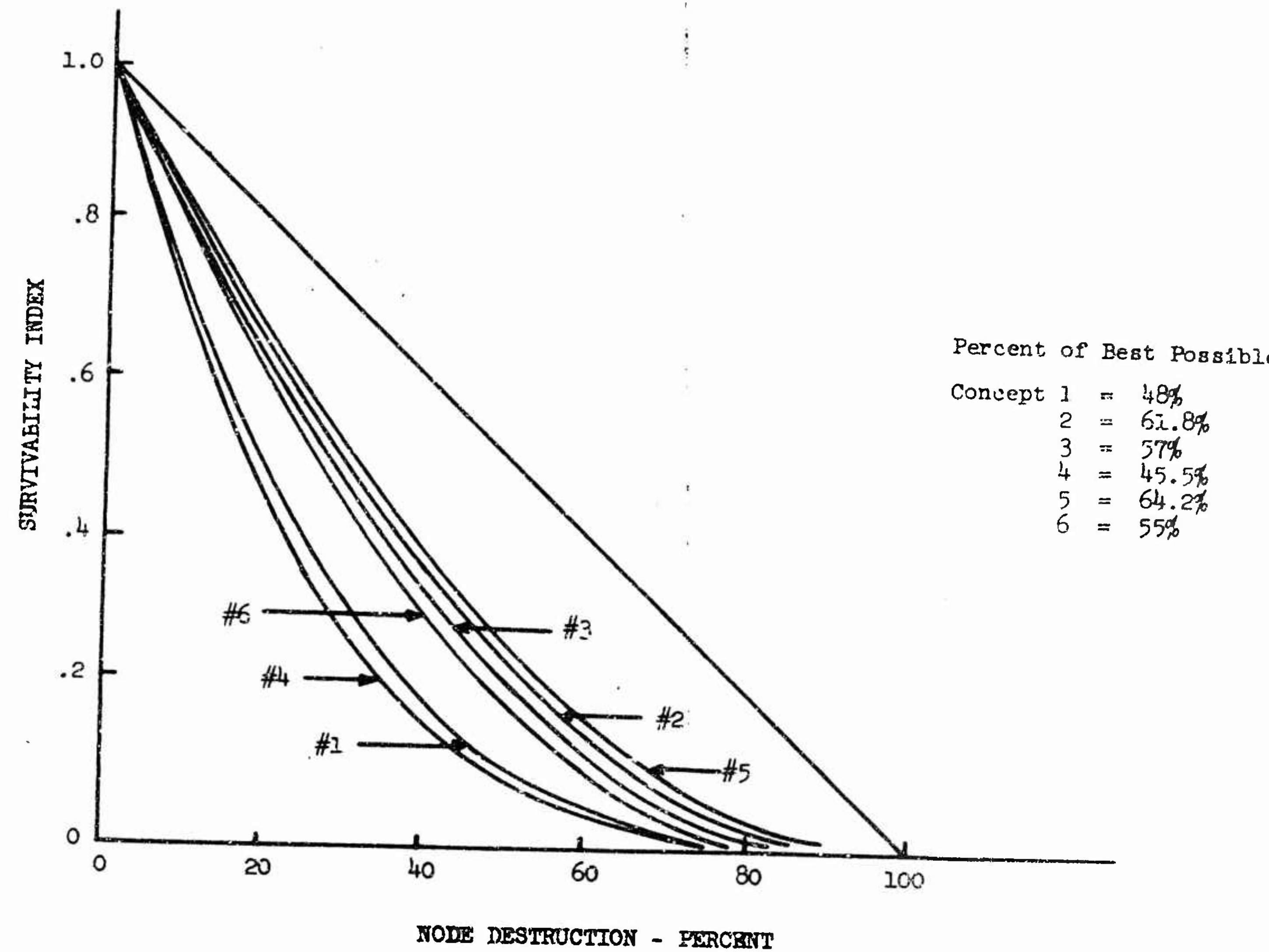


FIGURE 12 CONCEPT COMPARISON SURVIVABILITY CURVES

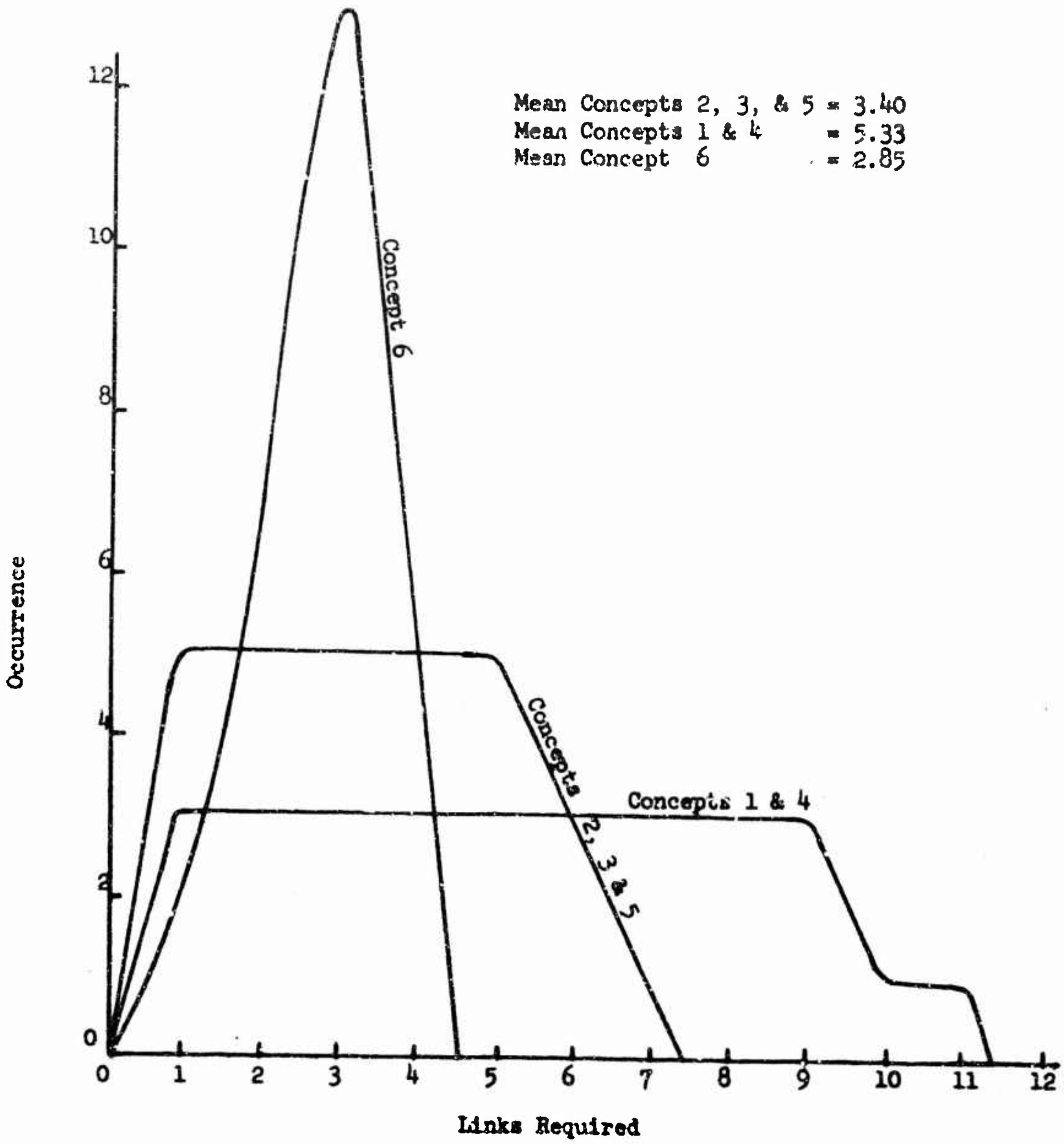


Fig. 13 Histogram of Links Required to Reach all Nodes From the Center Field Army Access Node

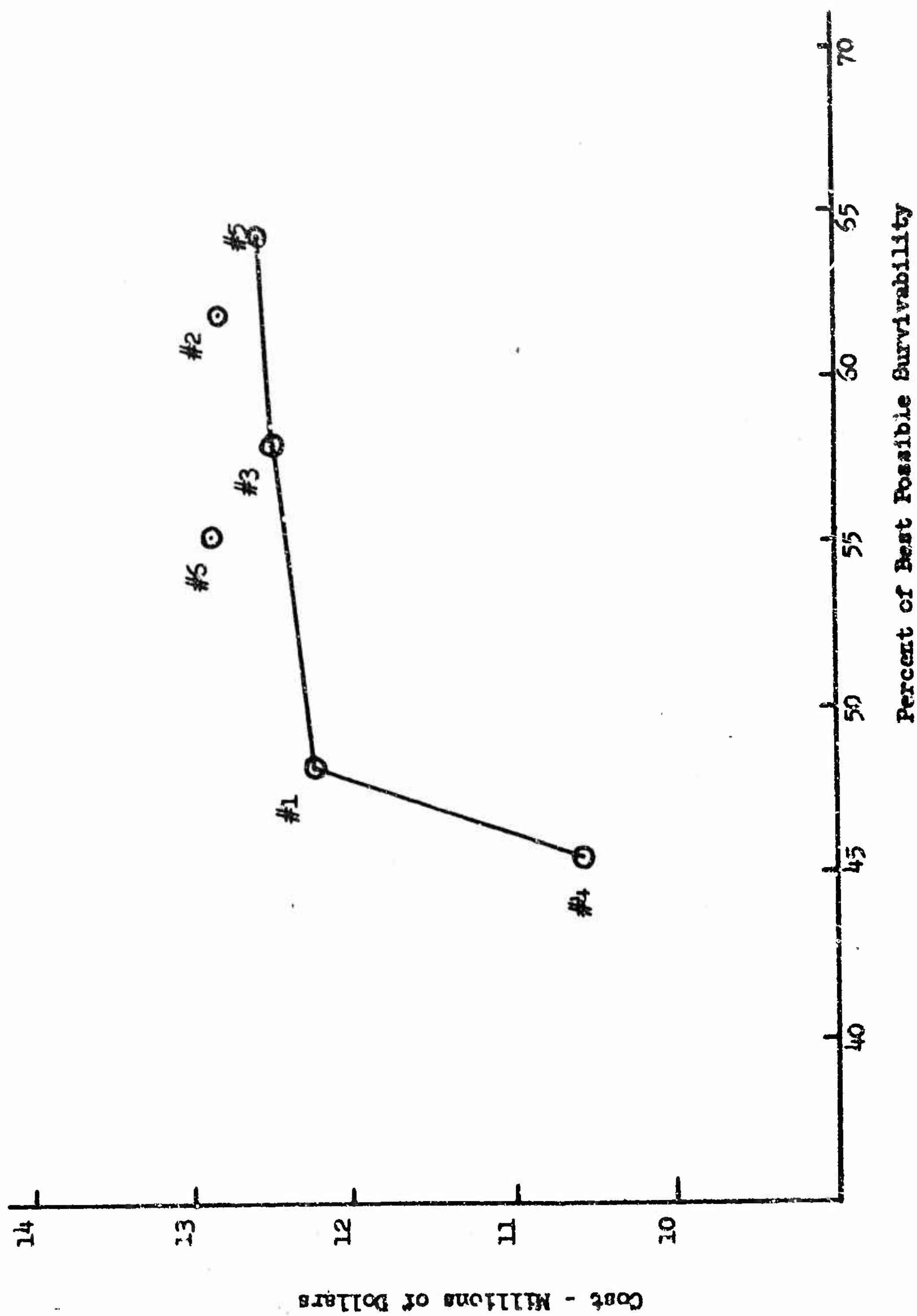


Fig. 14 Cost - Effectiveness Curve Initial Cost

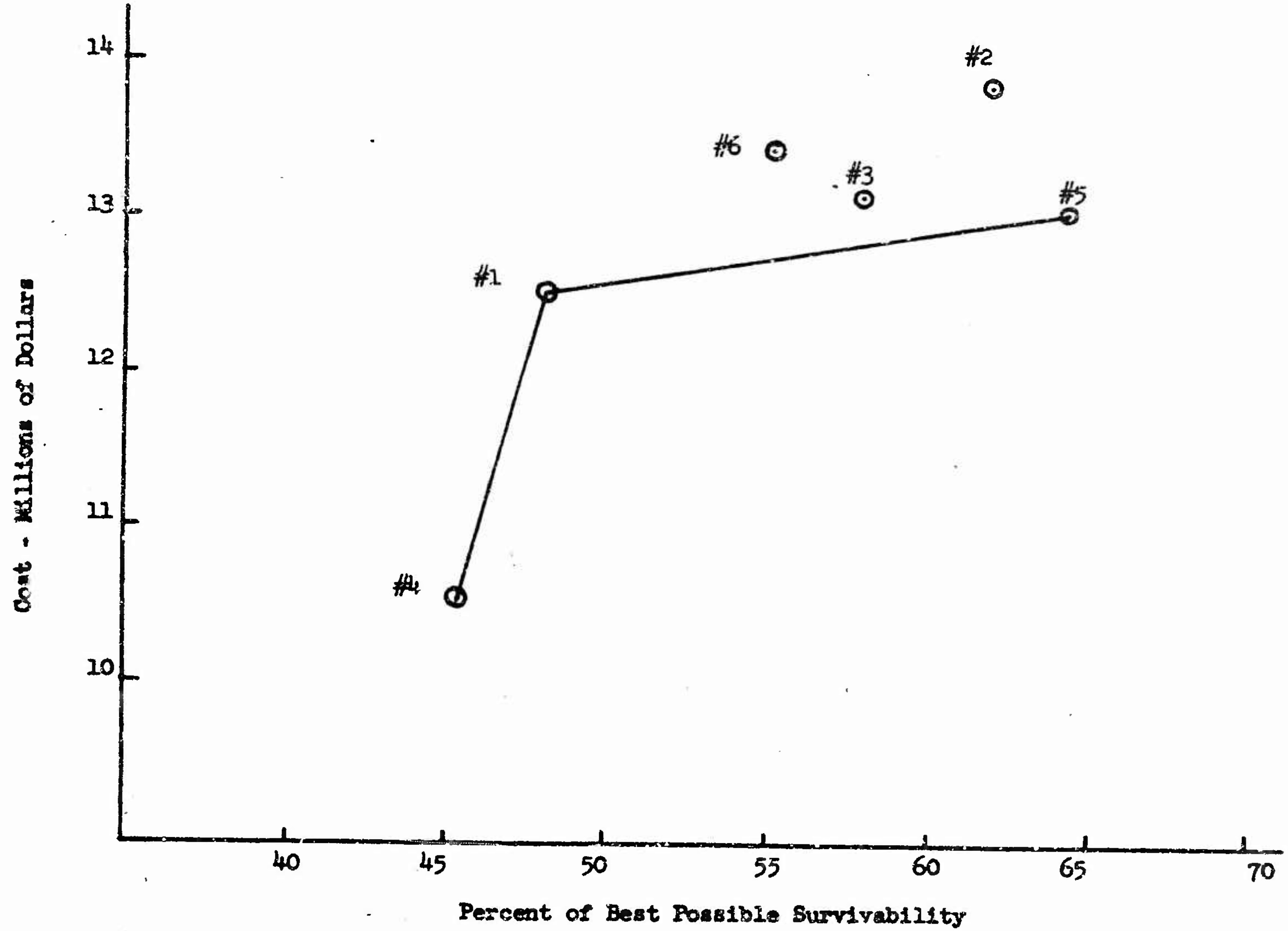


Fig. 15 Cost - Effectiveness. Curve Initial Cost Plus O & M Costs for One Year

A MODEL FOR DETERMINING TARGET LOCATION ACCURACY REQUIREMENTS

by

Mr. Otis S. Spears
U. S. Army Combat Developments Command
Artillery Agency
Fort Sill, Oklahoma

1. PROBLEM: To develop a methodology for determining the accuracy requirements of target acquisition systems.

Background remark: An important problem in weapon systems analysis is to determine the accuracy requirements of target acquisition equipment. Numerous models exist for overall weapons evaluation, but the target location accuracy requirements are usually assumed or estimated from cursory methods of approximation. With the prominence which target acquisition has recently assumed, emphasis on methods for determining accuracy requirements is timely. In fact, the accuracy requirements may become determining factors in selection of a target acquisition system from a set of competitive proposed systems.

2. DEFINITION OF TERMS: Certain terms peculiar to this study, or to this study area, are defined below:

a. Effects Pattern: The average area within which damage to enemy targets can occur as a result of exploding missiles or cannon volleys. In the current study, only fragmenting projectiles are considered, and the effects patterns are assumed to be circular.

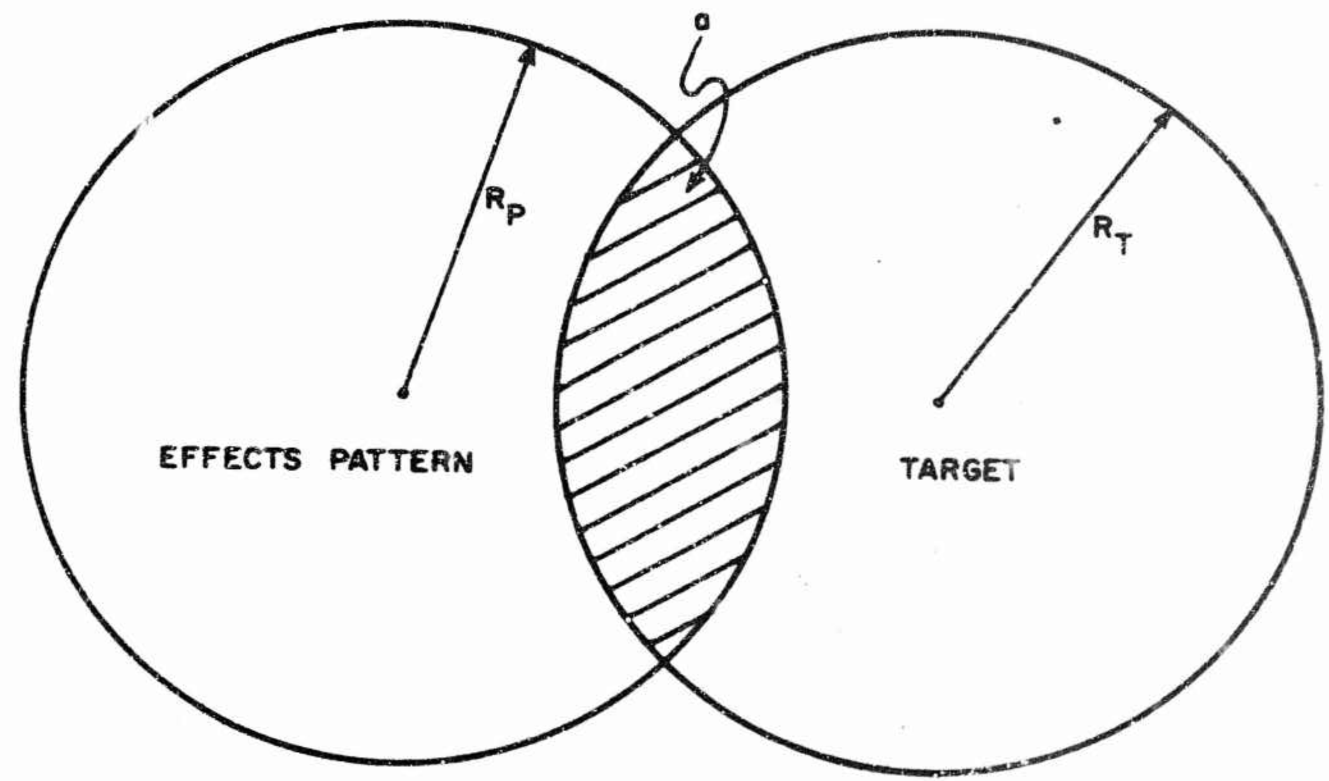
b. System Error: A measure of the distribution of effects pattern centers around the mean; usually expressed as a probable error or circular probable error.

c. Target: A specified enemy unit, installation, or area which is to be engaged. In this study, targets are assumed to be circular. This is for convenience, not necessity.

d. Coverage: The area common to the target and the effects pattern. This is sometimes referred to as "target coverage," or "area of coverage." The concept is illustrated by Figure 1. In the figure, the shaded area "a" represents the coverage. The ratio of "a" to the area of the target is called "fractional coverage of the target."

e. Effects: The fraction of casualties or fraction of damage on a given target, or within an effects pattern.

THE CONCEPT OF COVERAGE



R_P IS THE EFFECTS PATTERN RADIUS

R_T IS THE TARGET RADIUS

a IS THE AREA OF COVERAGE

FIGURE I

f. Lethal Area: A measure of casualty potential associated with a projectile bursting on or over a specified target. It (lethal area) is essentially the integral of a probability-of-casualty function, computed in the plane.

g. Target Location Error: Distance between a specified point within a target (usually the target center) and the supposed location developed from target acquisition information. This is a "systematic", or "bias", error for a given battlefield mission, at least when multiple volleys or missiles are fired at the supposed location.

3. ASSUMPTIONS:

a. All targets are circular.

b. All error distributions are circular "normal."

c. Effects patterns are circular, and the average casualty or damage level within each is uniform.

d. Horizontal components of vertical (fuzing) errors combine with horizontal system errors to form total (horizontal) system errors.

The assumptions listed above are for convenience. They are not necessary conditions for the basic theory, as will be seen below.

4. SCOPE AND APPROACH:

a. This study encompasses a method of determining target location accuracy requirements when other components of the weapon system are specified. The method is primarily developed for surface-to-surface artillery weapons, but is adaptable to other systems. The component errors assumed to be specified are as follows:

(1) The delivery system (e.g., a division general support missile).

(2) Survey.

(3) Meteorological errors.

b. Furthermore, the method is directed primarily at nonnuclear tactical situations, where the target acquisition problem is often acute. The procedures, however, can be adapted to nuclear weapons employed.

c. With specified circular effects patterns, along with other factors mentioned above such as target size and accuracy, one may study changes in single volley target coverage as the target location error changes. "Fractional coverage" is the fraction of target covered by the circular effects pattern. Thus, it is possible to study variation in the effectiveness of a weapon unit by observing the change in the fractional coverage associated with a target location error, without actually calculating the effectiveness. Moreover, it has been found that, for certain important cases, effectiveness realized with multi-volley or multi-warhead missions is directly proportional to single-volley fraction of coverage. This principle is very important, for it simplifies an intricate matter.

d. So, the change in effectiveness of a weapon system may be studied with respect to a simple quantity associated with the target location error: target coverage. Certain implications of these changes in effectiveness are relevant to the question of when reduced effectiveness becomes significant. Generally, reduced effectiveness becomes significant when the probability of accomplishing a mission is critically reduced from what it is when perfect target location is assumed.

e. For the present study, accomplishment of a mission is synonymous with defeat of an enemy target; but this does not necessarily imply an arbitrary damage level as a defeat criterion. Defeat of a target is assumed to occur when the enemy unit under fire "breaks"; that is, becomes ineffective. Damage levels are associated with corresponding probabilities that enemy units will break, or become disrupted for given time periods. Therefore, no specific defeat criterion (such as 30% casualties) is used. For example, an enemy unit dug in is not as likely to become ineffective with a given casualty level as an unprotected unit in the assault. At least, such is the indication of available data used. A portrayal of average casualty levels and associated probabilities that enemy units will break is shown by Figure 2. Actually, there is no known analytical method for calculating the exact level of damage required to defeat a target. As casualties mount, however, a critical level may be attained after which the affected unit is no longer able to accomplish its mission. This critical casualty level is based on unit casualties and materiel losses. There are also other less tangible factors such as troop experience, esprit, morale, and leadership. As shown by Figure 2, available data indicate that a 30% casualty level will usually so disrupt and disorganize an attacking unit as to make it ineffective for a considerable period of time. The question of the accuracy of such data, however, is not a consideration in the present study. There is a further discussion of this matter in Reference 2.

f. In summary, since coverage is directly proportional to effectiveness (damage level), reduction in single-volley coverage is associated with a corresponding reduction in effectiveness. In turn, this reduction is associated with a corresponding reduction in the probability of defeat. Thus, a fairly simple and important relationship is established between increasing target location errors and decreasing probabilities of accomplishing a mission. The increasing target location errors are first associated with decreasing single volley coverage, in order to complete the association with probability of defeat. Examples of the correspondence between reduced single volley coverage and reduced probability of defeat are shown in Figure 3. Information on the vertical axis (probability of accomplishing the mission) is based on Figure 2. The curves shown are based on different assumptions regarding the type of target. Furthermore, these particular curves are based on equal percentage reductions in effectiveness and single-volley coverage, and they show that reductions of the order of 10% are significant.

5. ANALYTICAL METHODOLOGY:

a. The lethal area of a weapon is a measure of casualty (damage) potential, usually measured for individual warheads or cannon rounds. Let g (x, y) denote the probability that a target centered at the point (x, y) will be a casualty from a fragmentation projectile which bursts at the origin (0,0). The symbol A_L usually represents the lethal area. Then:

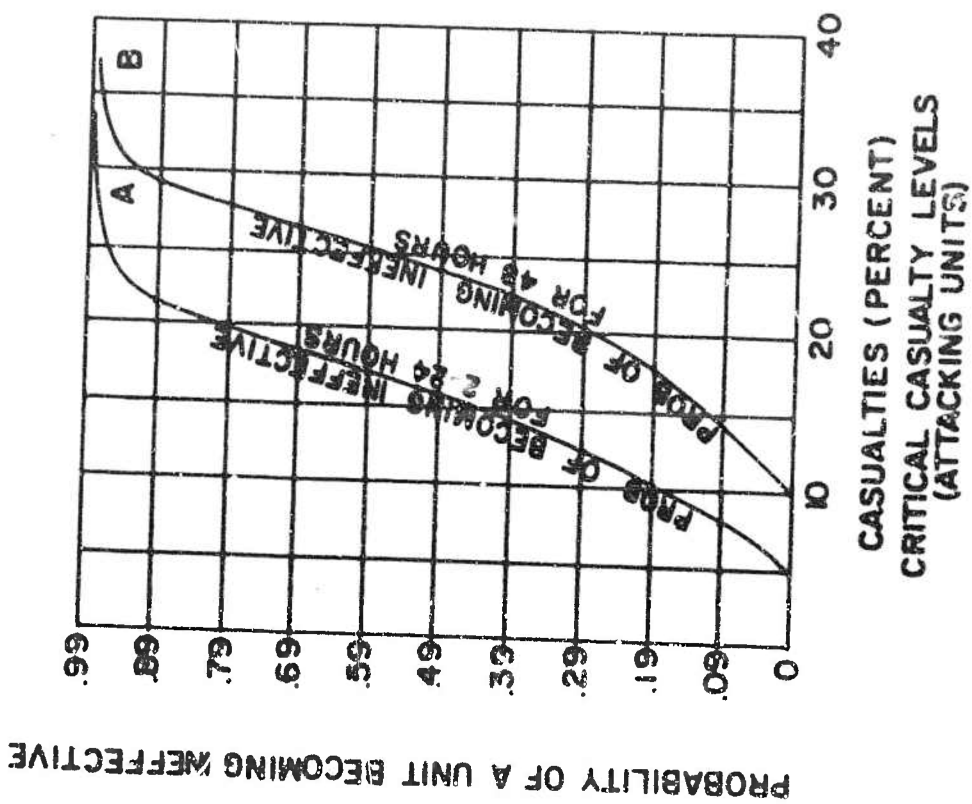
$$A_L = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g\,(x,\, y)\; dx\; dy \qquad (1)$$

Thus lethal area is not a specific geometric configuration, but rather a "summation" of casualty potential. This quantity, however, by the nature of equation (1) is expressed in terms of square units, such as square meters. Its magnitude is a function of the fragmentation pattern of the projectile, as well as the shielding factors applicable to the target (e.g., personnel standing, or personnel prone). Therefore, lethal areas are specified with respect to the "hardness," or shielding of the target. The quantity A_L, usually expressed in square meters, could theoretically have any non-negative value.

b. The fundamental unit of firepower considered in this study is the battery volley, or single missile. Let A_p be the area of the effects pattern. Then, the measure of effectiveness (f) within the unit pattern is as follows:

$$f = 1 - e^{-\frac{NA_L}{A_p}} \qquad (2)$$

FIGURE 2


A
B
PROB OF BECOMING INEFFECTIVE FOR 2-84 HOURS
PROB OF BECOMING INEFFECTIVE FOR 48 HOURS
.99
.89
.79
.69
.59
.49
.39
.29
.19
.09
0
10
20
30
40
CASUALTIES (PERCENT)
CRITICAL CASUALTY LEVELS
(ATTACKING UNITS)
PROBABILITY OF A UNIT BECOMING INEFFECTIVE

The quantity N is the number of projectiles in the volley. This is the so-called "cookie cutter" concept. Implicitly, the quantity A_L is treated here as a "solid area." This approximation results in no significant error when the value of A_L is small compared to the area of the effects pattern.

c. Let the total system error be denoted by CEP_s, and each component CEP be represented by E_i, where i designates the i-th component. Then,

$$CEP_s = \sqrt{\sum_i E_i^2} \tag{3}$$

The target size, the quantity f, the effects pattern size, and CEP_s determine the effectiveness, or fraction of damage. The fractional coverage associated with a given probability is then calculated analytically, or read from P(f) nomographs. (See Reference 4.) When read from nomographs, the determining numbers are the following ratios: R_p/R_T and CEP/R_T. R_p and R_T are the radii of the effects pattern and the target, respectively. Then, for a fractional coverage C', the fraction of casualties F is,

$$F = fC' \tag{4}$$

Thus, F is directly proportional to coverage for a fixed f. A fixed f within each effects pattern has been assumed (paragraph 3c). Equation 4, however, is for one volley only.

d. Next, a few more important quantities and their symbols are: Area of the target, A_T; fraction of casualties, or damage, F; the area of the effects pattern, A_p; the area common to the target and the effects pattern, a; the lethal area of the i-th projectile or warhead, A_{Li}; average fractional coverage of the effects pattern by the target, C; average fractional coverage of the target by the effects pattern C'. Thus, the fractional coverage C is,

$$C = \frac{a}{A_p} \tag{5}$$

Likewise,

$$C' = \frac{a}{A_T} \tag{6}$$

Figure 1 shows a representation of the target circle, the effects pattern, and the overlap area a.

e. When a series of projectiles or missiles is fired at a target, a measure of the total firepower is $\sum_i A_{Li}$. Even if all projectiles

fired are of the same type or caliber, the lethal area may be different for different projectiles (or missiles), because the target condition may change; e.g., personnel may change from a standing to a prone posture. A measure of the average fraction of the firepower which hits the target is $C \sum_i A_{Li}$. Then, the effectiveness F is,

$$F = 1 - e^{-\frac{C \sum A_{Li}}{A_T}} \tag{7}$$

Now, let,

$$k_1 = \frac{A_T}{A_p} \; ; \; k_2 = \frac{\sum A_{Li}}{A_T} \tag{8}$$

Also, let,

$$k_3 = k_1 \, k_2 \tag{9}$$

Therefore,

$$k_3 = \frac{\sum A_{Li}}{A_p} \tag{10}$$

From (7) - (10), inclusive,

$$F = 1 - e^{-C'k_3} \tag{11}$$

Equation (11) exhibits the effectiveness F in terms of the following important quantities: The portion of the target that is covered by the effects pattern (C'), and the ratio of the total firepower to the effects pattern (k_3).

f. As already explained, the essential principle of this method rests on the possibility of making inferences about reduced effectiveness from easily determined single-volley coverage of the target. In other words, the principle is that

$$\frac{dC'}{C'} = \frac{K dF}{F} \tag{12}$$

where K is either a proportionality constant, or a parameter. From equation (11), the principle is,

$$\frac{K \, dF}{1-e^{-C'k_3}} = \frac{dF}{C'k_3 e^{-C'k_3}} \tag{13}$$

This means that,

$$K = \frac{1-e^{-C'k_3}}{C'k_3e^{-C'k_3}} \tag{14}$$

For the particular case where K = 1,

$$e^{-C'k_3}\,(C'k_3 + 1) = 1. \tag{15}$$

Two facts are immediately evident from (14) and (15). The first is that K is not constant, but that it changes slowly for considerable changes in fractional values of C' and k_3. Secondly, it is also obvious that, for certain intervals, equation (15) is correct to a high degree of approximation. Figure 3 shows that changes in C' of 10% are serious under conditions of equation (15).

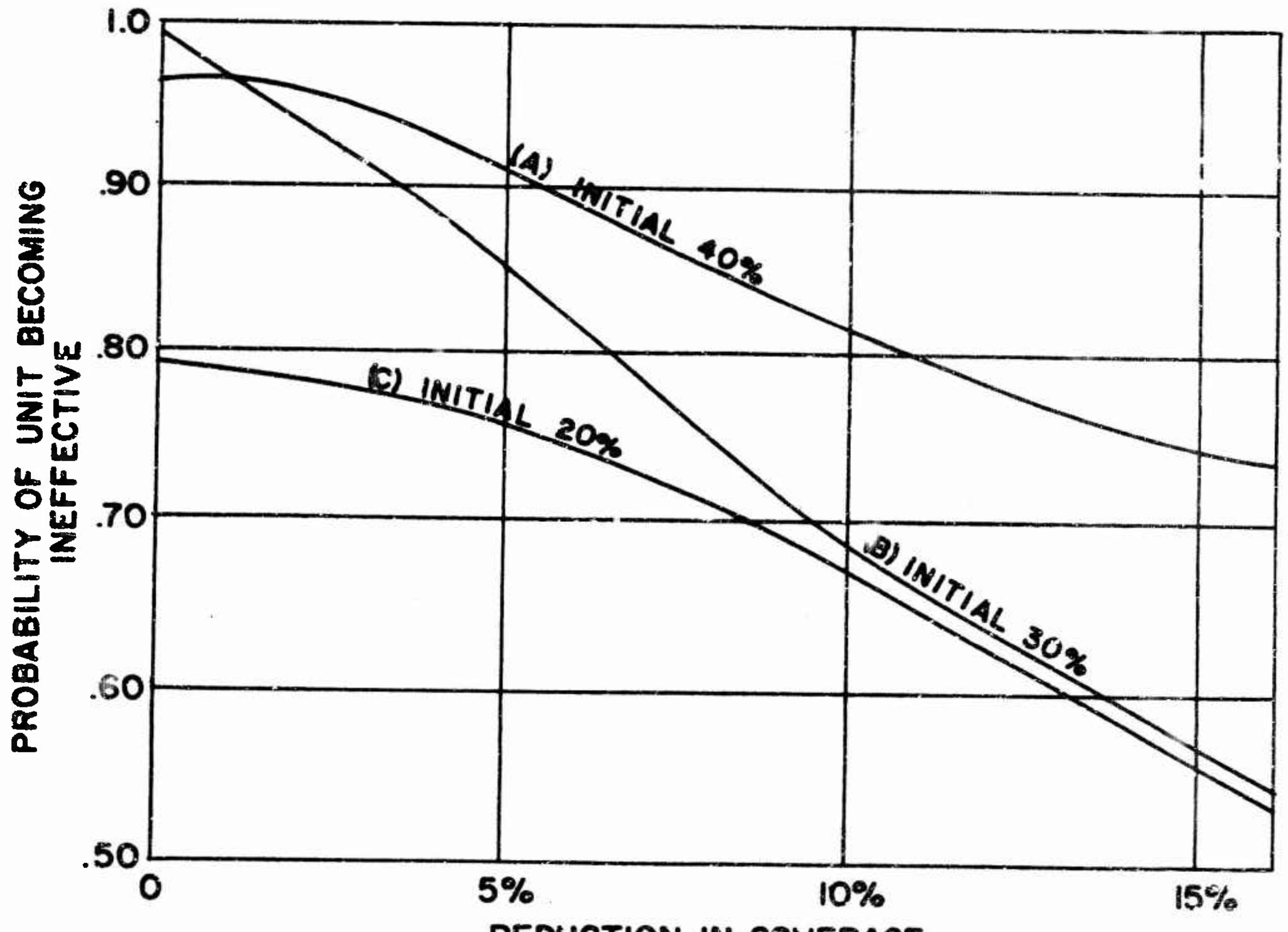
Table 1 shows that K is relatively stable for such changes in C', regardless of whether condition (15) holds. The upper portion of the table includes a total range typical of ordinary non-nuclear systems. The bottom portion includes certain extreme values, such as k_3 = 1.00.

TABLE I

THE STABILITY OF K UNDER REDUCTIONS IN C'

C'	k_3	$C'k_3$	$e^{-C'k_3}$	$1-e^{-C'k_3}$	$K = \frac{1-e^{-C'k_3}}{C'k_3e^{-C'k_3}}$
0.50	0.10	0.050	0.951	0.049	1.03
0.45	0.10	0.045	0.956	0.044	1.02
0.50	0.20	0.100	0.905	0.095	1.05
0.45	0.20	0.090	0.914	0.086	1.04
0.43	0.20	0.086	0.917	0.083	1.05
0.75	0.50	0.375	0.691	0.309	1.19
0.64	0.50	0.320	0.726	0.274	1.18
0.70	0.50	0.350	0.705	0.295	1.19
0.60	0.50	0.300	0.741	0.259	1.16
0.70	1.00	0.700	0.497	0.503	1.45
0.63	1.00	0.630	0.533	0.467	1.39

g. Finally, a simulation study was performed, for certain cannon units, to check Table 1. Multiple volleys (from 2 to 5) were used.



CHANGE IN PROBABILITY OF EFFECTIVENESS
vs.
REDUCTION IN COVERAGE

FIGURE 3

Coverage and effectiveness were measured empirically. Changes in C' and F were then measured. Similar computations were then made from equation (11) and the data in Table 1. Essentially, the entire exercise, including the simulation based on experimental firing data, was to compare the predicted and empirical values of $\Delta C'/C'$ and $\Delta F/F$. Results are shown in Figure 4. The two sets of values show almost perfect agreement. It is emphasized that these results are based largely on typical values of C' and k_3. For example, single-volley coverage (C') of more than 70% of a target would be an excellent achievement in most battlefield situations.

6. CONCLUSION: Changes in single-volley coverage of a target by a weapon effects pattern (a quantity relatively easy to determine) can be used as a basis for determining critical reductions in effectiveness of multi-volley fire (a quantity difficult to determine accurately).

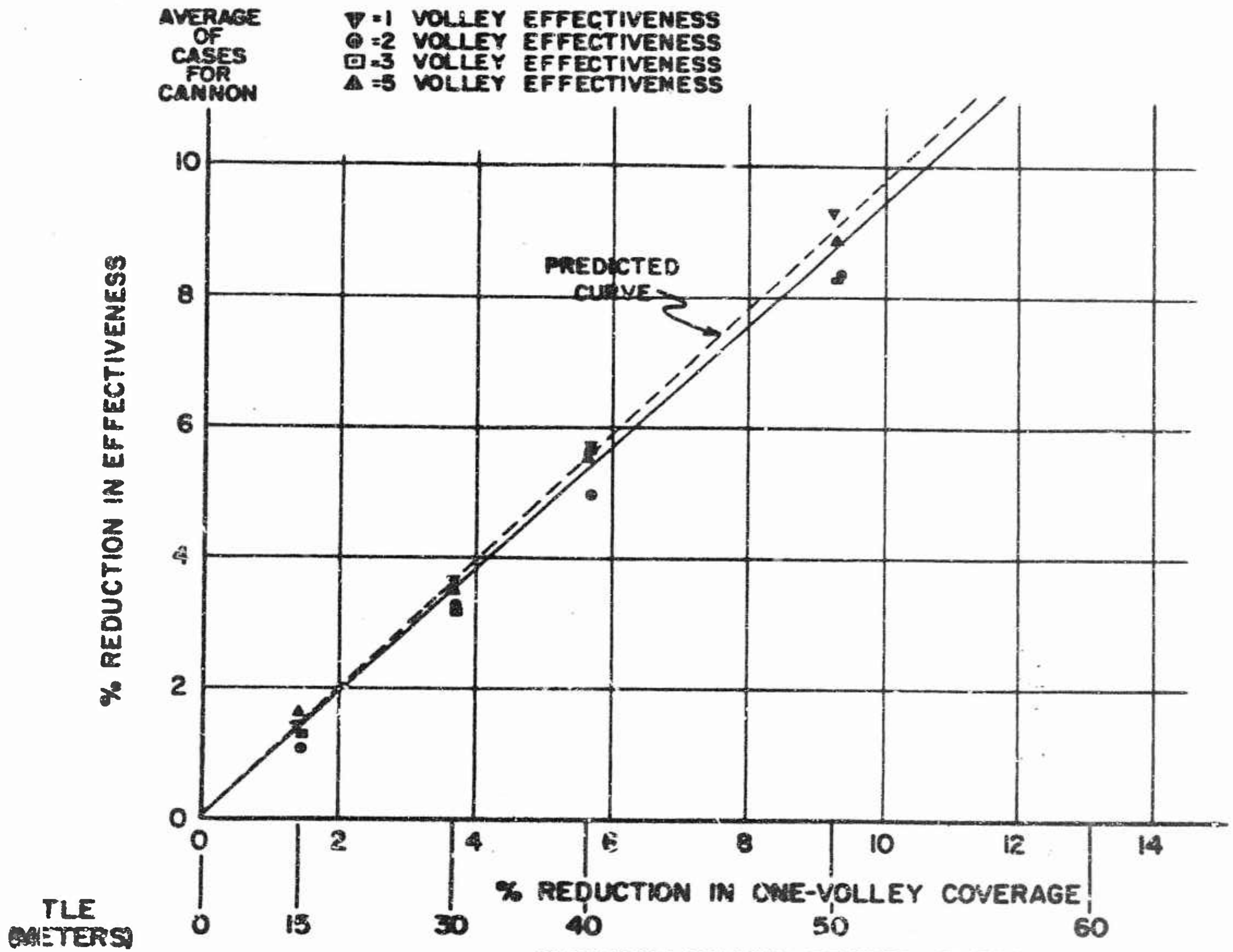
REFERENCES

1. Study, "Target Acquisition Accuracy Requirements 1965-75 (U)," USACDC Artillery Agency, Fort Sill, Oklahoma, June 1965.

2. Field Manual 6-141-1, "Nonnuclear Employment of Field Artillery Weapons Systems (U)," Department of the Army, September 1964. (Manual is classified CONFIDENTIAL.)

3. Handbook, "Probability and Statistics with Tables," by Burington and May, published by Handbook Publishers, Inc., Sandusky, Ohio, 1953.

4. Staff Officers Field Manual 101-31, "Nuclear Weapons Employment," Hq, Department of the Army, February 1963.



COVERAGE VERSUS EFFECTIVENESS
155 BATTERY HE

FIGURE 4

USING OPERATIONS RESEARCH TO EXPLORE DESIGN CONCEPTS FOR A NEW UNIT RECORD

BY

Mr. Donald D. Curry
Electronic Engineer
U. S. Army Electronics Command

Mr. Donald F. Blumberg
Manager, Operations Research and
Long Range Planning Department
TECHREP Division, Philco Corp.

ABSTRACT

Operations Research was used to determine a Unit Record specification which is designed to meet present and future military user requirements and overcome the inherent deficiencies of the present unit record system, the punch paper card. The new Unit Record has inherent to it higher data density, provision for imaging, and authentication, all with utmost flexibility for any application along with maximum consideration for human handling of the Unit Record medium.

I THE UNIT RECORD PROBLEM

The rapid expansion of military and commercial data processing maintains a constant demand for continuous improvement of computer systems and input/output equipment. This demand has been met in recent years with faster nd improved circuits, core, and film memories, disks, tapes, drums, etc. In unit record punched card systems, however, little has been accomplished to meet the requirements of the newer and faster computer systems. Less has been accomplished for militarized, low power, weight and size unit record devices and no economical solution has yet been offered to solve the problem of militarized unit record materials.

As a clarifying point, a unit record is defined as a separable document which contains information about one item and is both man and machine readable. Separable document is taken to mean a document that can be removed from and put back into the same file containing other similar documents.

For many years now it has become increasingly apparent that there is a major discrepancy between the efficiency of unit records as the input/output functional or source documents, and the efficiency and use of a high speed central data processor.

One form of these source data documents is the punched card. This punched card is one of the most important man-machine interfaces in today's information processing systems, but this punched card as an input for Automatic Data Processing Systems is becoming obsolete. These cards are reaching the limit of their performance capabilities. Some of these limitations are:

1. Punched cards do not satisfactorily handle multi-media. Diverse forms of data such as digital, man-readable a/n characters, pictorial, and signature validation can not be met by present unit record methods.

2. There is the inherent limitation of low information density on the present unit record. Many applications have unit records processing several hundred characters, thus demanding many trailer cards for each unit record. As the Army's information processing is automated even more in the future, more and mo trailer cards will be demanded for unit record storage because automated system procedures demand more data.

3. The present unit record requires expensive, cumbersome, and unreliable equipment for input/output of data.

4. This present equipment operates at a speed not compatible with modern machine processing speeds. The present unit record speed is considerably slower than the speed of a modern data processor.

5. And finally, the present unit record is totally unable to meet the stringent combat field environment conditions that exist.

The present lack of available data capacity to describe one item and very slow "processing time" certainly points the way toward a definite need for replacing the present ADP punched paper card. The military services in a military field combat environment certainly need a more sophisticated form of a Unit Record, one that would meet the stringent requirements of combat and provide a higher character data density in addition to imaging and legal considerations. There obviously exists a need for a new, technologically up-to-date UNIT RECORD.

II METHOD OF ATTACK ON THE UNIT RECORD DESIGN PROBLEM

The normal engineering approach to systems design as related to unit records is depicted in Figure 1, "Standard Engineering Approach to Systems Design". One would concurrently investigate present and future processing technology available along separate paths. After the best material and the optimum processing was selected, then these two would be combined into the Operational System Design from which would evolve the Design Specification. That is the conventional approach.

STANDARD ENGINEERING APPROACH TO SYSTEMS DESIGN

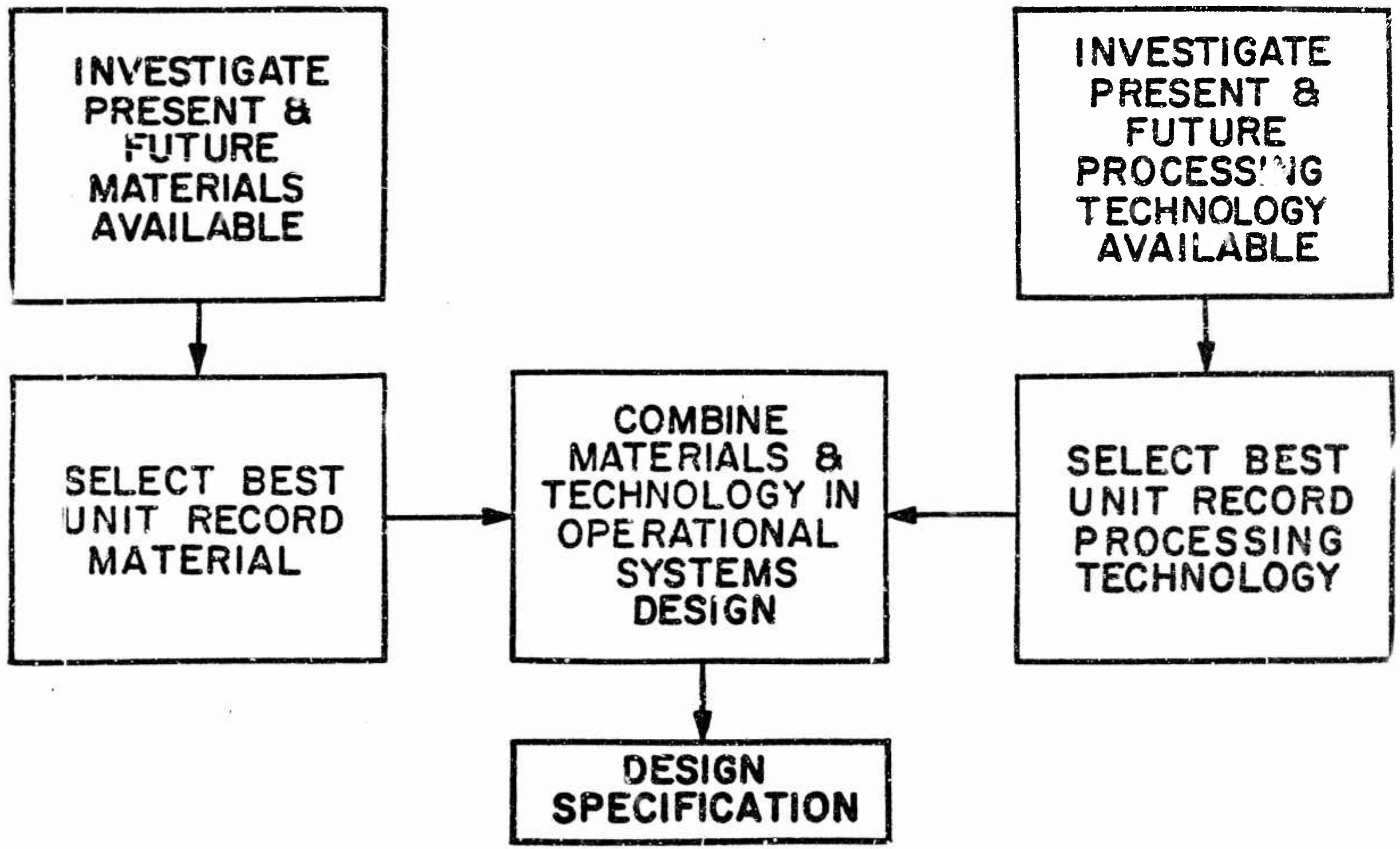


FIGURE I

There are a number of things wrong with this type of approach:

1. It doesn't benefit user potential.
2. It doesn't force a cost-benefits analysis.
3. It commits to a specific design at too early a stage.
4. It doesn't match ultimate consumer requirements (in operational environment) with design specifications.

For the above reasons an Operations Research approach was used to determine the final systems design of the Unit Record concept to achieve an optimum design. Approaches other than an Operations Research approach generally do not produce such an optimum design. An operations research approach depicted in Figure 2, "Use of Operations Research in Approaching Systems Design", was used to the fullest extent possible. This offers the Army the maximum benefit of studying the potential user and planning to meet his future needs. It is pointed out in this paper how a Unit Record Specification was determined through Operations Research. The following Operations Research approach was used.

First an "Operational Analysis of User Needs" was made currently with an "Analysis of Parametric Constraints". This, through a "Simulation of Alternative Systems Concepts to meet User Needs" determined our "Definitions of Systems Specifications in Terms of User Needs". From here an "Assessment of Specifications" was made in light of the presently available and future creation of materials and technology. This leads to a "Cost-Benefits Analysis of Design Alternatives" from which is achieved from the Operations Research Approach the "Final System Design". This Operations Research Approach most thoroughly achieves a farsighted and competent design.

Starting with the Analysis of User Needs, a survey was made of the fields that have the greatest potential for future use of Unit Records: Personnel, Logistics, Military Police, and Command and Control. Also, the specific functions and data required to perform their various and diverse functions were studied.

From just a brief look at the functions of the various fields studied, it can be seen that a new unit record needs to handle several diverse forms of data. Therefore the functional requirements of a new unit record can be divided into four functional areas:

1. Machine Readable Data
2. Man Readable Information
3. Image Storage
4. Signature or Authentication

Each application has various requirements for different functional storage.

USE OF OPERATIONS RESEARCH IN APPROACHING SYSTEMS DESIGN

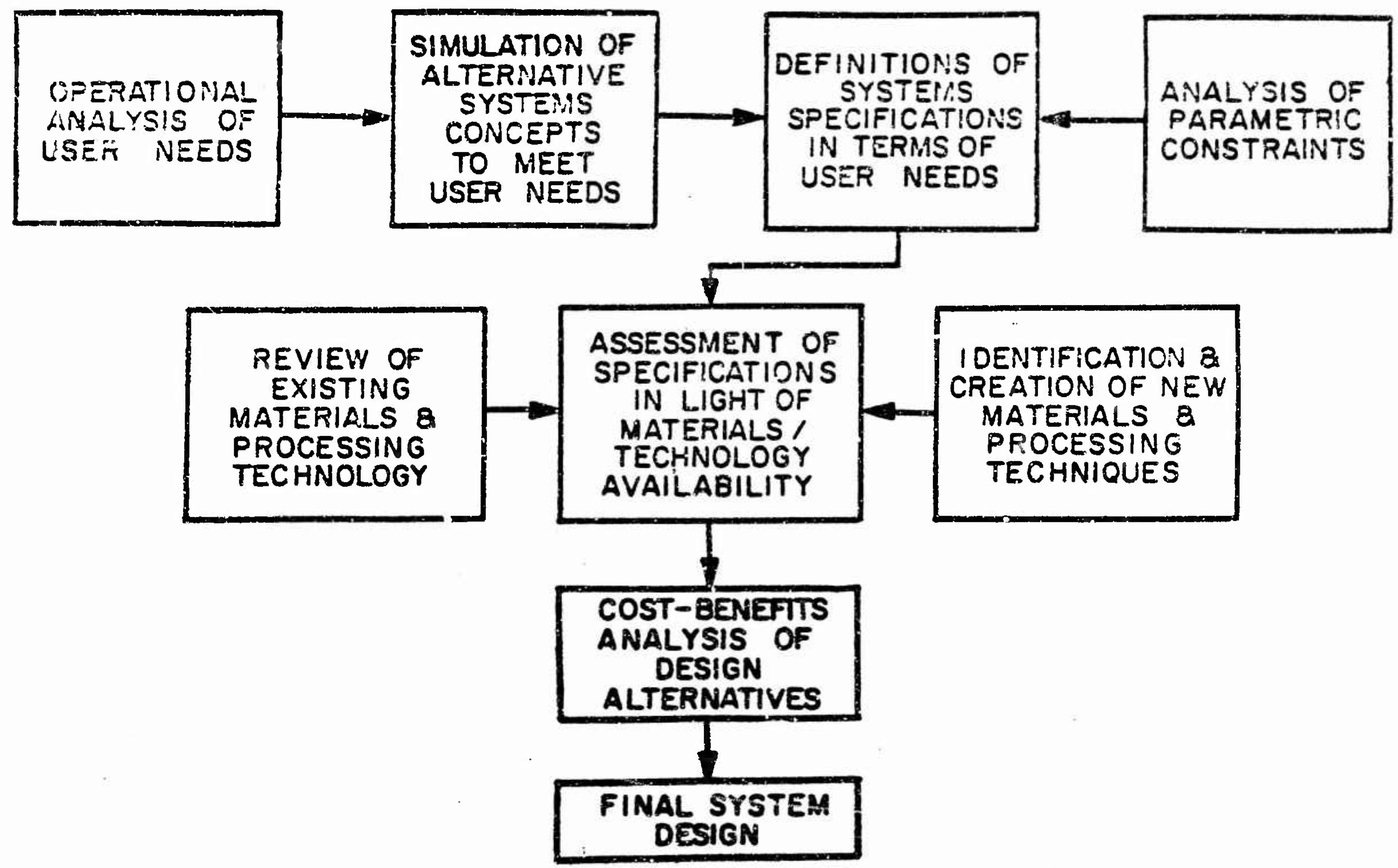


FIGURE 2

In studying Personnel applications a survey was made of the personnel and administrative functional requirements of records keeping which are basic to the strength and accounting systems, assignment of replacements, rotation of personnel and reporting of casualties. This was accomplished by holding discussions with representatives from the Adjutant General Agency, Fort Benjamin Harrison, Indiana, and studying applicable documents. These personnel functions include the initiation, maintenance, use and disposition of data information required in the execution of personnel management functions within the army in the field, from the unit level to the theater of operations. It would be most beneficial to the army to reduce the manual initiation and maintenance of such records involved in the above functions to a minimum, especially at the small unit level. Also to permit the timely accumulation and distribution of individual and/or unit personnel statistical data for use in the performance of personnel management and manpower control functions in the army in the field. A system must be developed, consistent with communications and ADP capabilities, which will promote flexibility of operation with optimum effectiveness in the control of human resources. Whatever Unit Record concept is chosen for personnel record keeping must be designed to serve the supported commander by the maintenance of a chronological, historical record to be utilized in the exercise of command in the performance of personnel management and manpower control responsibilities.

One aspect of Logistics was Ordnance Ammunition Supply. Talks were held with Logistics oriented people responsible for establishing the inventory control requirements for the future, 1965-1970. The results of these discussions were an understanding of a future comprehensive Logistics supply system to effectively utilize automatic data processing capabilities, integrate total available assets to best support total requirements, extend the current army ammunition credit system down to division level organizations, and to centralize control at theater Army level and field Army level. Inventory control includes managing, cataloging, requirements determination, procurement, distribution, overhaul, and disposal of ammunition. Additional functions include supply and stock control, and stock management, as well as the records and reports required to effectively and efficiently accomplish these functions. Unit Records would serve most useful if located at each ammunition storage installation for the purpose of providing receipt, issue, change of condition, and inventory adjustment information to the appropriate inventory control center.

Discussions held with the Military Police planning and concepts people at Fort Gordon, Georgia, made it possible to obtain invaluable information on the Military Police functional areas of law enforcement, police traffic management, and confinement administration. Law enforcement embraces the recording, reporting, and management tasks associated with present military police reports, MP criminal investigation reports, MP traffic accident reports, and Armed Forces traffic tickets. Police traffic management includes

the registration of privately owned vehicles, the maintenance of driver violation records, and the maintenance of source data on traffic accidents needed for analysis. Confinement administration includes all those recording and managerial tasks for stockades, hospital prisoner wards, and rehabilitation training centers. While designed for the army in the field, the proposed Unit Record system is also applicable in CONUS posts, camps, and stations.

Turning to the Military Police discipline, law, and order functions, a Unit Record concept would certainly provide unique capabilities of automation and offer a great potential for the more rapid detection and apprehension of offenders. Military Police requirements for a Unit Record information system may be described by two prime needs. First, there must be mass data storage and ready retrieval of distilled information or retrieval of the basic input medium itself as needed. Second, there must be a flexible system for obtaining this offense information for analysis and review. The size of the problem may be seen in considering the police offense reporting requirements for a theater army. Of an estimated 65,000 police reports generated per month in a theater army of 1.5 million men, at least one-half will require some type of collating search.

An automated prisoner of war record keeping and reporting system is certainly warranted because of the present extensive and time consuming clerical effort involved. It is proposed that a Unit Record concept certainly should meet the needs of this function. A Unit Record concept should be directed toward those prisoners of war record keeping and reporting requirements which come into effect with their administrative processing in the communications zone and the preparation of individual personnel records. Requirements for the maintenance and updating of such records and for the submission of reports would continue throughout the period during which the prisoners of war are held in United States Army custody. If unit records were used this would greatly minimize the number of forms and records which must be manually prepared and transmitted or retained in individual prisoner of war files, and would provide statistical data and information in such a manner as to be fully responsive to the requirements of the commander. With the processing and establishment of individual records for each PW, the clerical burden imposed assumes very significant proportions. Under the present manual system, the major portion of this burden rests with the PW Camps. A Unit Record concept should be directed toward the development of automatic PW record keeping and reporting system which will relieve the PW camp commanders of the burden to the maximum extent possible and, coincident therewith, provide PW data which will be fully responsive to command requirements.

It can be seen in Figure 3, "Operational Analysis of User Needs", that Personnel applications will require a high degree of machine readable data storage, while Logistics applications will require a high degree of man readable information storage, and signature authentication. If the Unit Record concept is to meet the needs of the Military Police, it will need a

OPERATIONAL ANALYSIS OF USER NEEDS

APPLICATIONS	REQUIREMENTS OF UNIT RECORD			
	MACHINE READABLE DATA STORAGE	MAN READABLE INFORMATION STORAGE	IMAGE STORAGE	SIGNATURE
PERSONNEL	*	●	●	●
LOGISTICS	●	*	●	*
MILITARY POLICE	*	●	●	●
COMMAND & CONTROL	*	▲	*	▲

NOTATION:
HIGH —— *
MEDIUM- ●
LOW —— ▲

FIGURE 3

high degree of machine readable data storage. Command and Control applications include intelligence and tactical operations functions. The high image storage requirement here is taken to mean displays. By looking deeper into each of these applications, a systems concept was arrived at to show the probability of search, the average record length, and the probability that the record did not exceed certain record lengths. See Figure 4, "Alternative System Concepts". By analyzing various records kept at each of our studied applications, we were able to determine the probability of the file being certain lengths. From this analysis it was determined that a record length on the order of 1000 characters was quite reasonable and responsive to the studied applications. The probability of each shows what percentage of records need to be searched on each transaction.

Based upon and time phased with this operational analysis of user needs was the Analysis of Parametric Constraints. These are listed on Figure 5, "Parametric Design Factors Required in Army Unit Record Concept". The first of these factors, and one of the most important, is "Reliable and Operable in a Military Field Environment". Ideally this means the elimination of critical mechanical adjustments and, of course, implies operation under the most severe environmental conditions. Next in importance is manual operation of the Unit Records under emergency conditions. This manual operation has two meanings and implies dual environment. One as stated where emergency conditions are defined as a complete absence of any power sources and would allow operation by only equipment that can be handcarried. The other is manual operation under full power where operations occur similar to present operations or punch cards. The third factor, "Low Cost Concept", is important regardless whether the Unit Records are disposable or reuseable. For either approach to be considered for the final concept, both need to be extremely inexpensive. "Ease of Preparation" is taken to mean that to prepare a unit record the techniques involved shall be no more complex than comparable present day punched card procedures. As seen later, the physical size of the Unit Record medium for manual handling as well as machine searching is also very important. Semi-Automatic filing is taken to mean a provision for filing the unit record in a cartridge or magazine or other similar mechanism. It is paramount that a group of separate but related records can be kept together and form a complete file on one topic.

Of course in studying and reviewing the existing and future material and processing technology, an extensive analysis was made, but for purposes relating to this paper the different diverse approaches were simplified to these three.

1. Eliminate unit records and use keyboard and/or voice input of source data to a computer and centralize filing in mass secondard memories.

2. Typewritten or hand-printed documents optically scanned locally and electrically transmitted or hand carried to central scanning points for direct input to mass memory and/or computer with local filing of the original documents at their source.

ALTERNATIVE SYSTEM CONCEPTS

APPLICATIONS	PROBABILITY OF SEARCH	AVERAGE RECORD LENGTH a/n	PROBABILITY THAT RECORD DID NOT EXCEED			
			100 a/n	500 a/n	1000 a/n	1500 a/n
PERSONNEL	.25	1300	.05	.10	.45	.80
LOGISTICS	.25	700	.05	.30	.65	.95
MILITARY POLICE	.50	1000	.05	.20	.50	.75

FIGURE 4

PARAMETRIC DESIGN FACTORS REQUIRED IN ARMY UNIT RECORD CONCEPT

FACTORS	WEIGHT (%)
• RELIABLE & OPERABLE IN MILITARY FIELD ENVIRONMENT	20
• MANUAL OPERATION UNDER EMERGENCY CONDITIONS	15
• LOW COST CONCEPT	8
• EASE OF PREPARATION	10
• SEPARABLE DOCUMENT	5
• CONVENIENT SIZE OF UNIT RECORD	12
• EASE OF USE BY UNSKILLED PERSONNEL	10
• FLEXIBILITY OF USE	8
• SPEED OF OPERATION	7
• SEMI-AUTOMATIC FILING CAPABILITY	5
TOTAL	100 %

FIGURE 5

3. A new high density separate Unit Record system with increased record and read speeds, smaller file sizes, small militarized card handling devices, and a storage medium suitable for military environments. The system should integrate with a manual document system in event of transceiver and/or computer failures.

At this point of juncture, it was not apparent which approach to follow. In fact , without following an Operations Research analysis of the problem these three different approaches would not have come to light.

It was needed to choose between these three approaches the optimum one to meet the Operational Analysis to follow. Therefore in assessing these approaches to parameters stated previously a decision table was developed (Figure 6, "Assessment of Approach to Constraints").

In this table the approaches were evaluated and assessed against the parametric constraints mentioned previously. The approaches were weighted by a numerical factor, scale 1 to 5. Certain parametric constraints lend themselves more to one of the three approaches. This readily can be seen in the manual operation constraint if one remembers manual operation means operation with lack of power. For the "Keyboard and/or Voice Input" approach there can be relatively no manual operation. While for the "Documents Optically Scanned" approach, it is possible to have manual operation. But the most desirable manual operation and most reliable would occur with the "High Density Unit Record" approach. Looking to another constraint, the ease of preparation, it is readily apparent that voice input would be the easiest for preparation since only voice is involved, while for any other approach the unit record requires a higher degree of preparation than voice input. Flexibility as a constraint is ideally suited to a unit record concept. Here with the four different functional areas: Man Readable Area, Machine Readable Area, Image Area, and Authentication Area, the Unit Record can be tailored to any application. The Unit Record technique developed should offer maximum flexibility for different classes of users. Certain users would require all digital storage, others would require all image storage, a third class could require combination of all four functional formats. Thus the Unit Record achieves utmost flexibility as compared to the rather limited and negligible flexibility of a keyboard, voice, and/or physical documents being optically scanned. These assessments were then multiplied by the particular parametric constraint weight and totaled. These final totals are 223, 302, and 440. The large total of 440 indicates Unit Records is the optimum approach. Now to establish the further specifications for the Unit Record concept.

Following the Operations Research approach outlined earlier, the next step is a Cost Benefits Analysis. One of the most important design alternatives is one of cost. Therefore, a Cost Benefit Analysis was made and plotted on Figure 7, "Cost of Updating Individual Unit Record". In this figure are shown two curves, one for reuseability, one for nonreuseability. This

ASSESSMENT OF APPROACH TO CONSTRAINTS

PARAMETRIC CONSTRAINTS (SHOWING WEIGHT)		APPROACHES		
		KEYBOARD AND/OR VOICE INPUT	DOCUMENTS OPTICALLY SCANNED	HIGH DENSITY UNIT RECORD
OPERATION IN FIELD ENVIRONMENT	(20)	3	3	5
MANUAL OPERATION	(15)	1	3	5
LOW COST	(8)	2	1	3
EASE OF PREPARATION	(10)	5	5	4
SEPARABLE DOCUMENT	(5)	1	3	5
CONVENIENT SIZE	(12)	1	2	4
EASE OF USE	(10)	3	5	4
FLEXIBILITY	(8)	2	3	5
SPEED	(7)	2	3	4
FILING	(5)	1	1	4
TOTAL ASSESSMENT (W TIMES A)		223	302	440

ASSESSMENT: 5 (EXCELLENT) ⟶ 1 (POOR)

FIGURE 6

COST OF UPDATING INDIVIDUAL UNIT RECORD

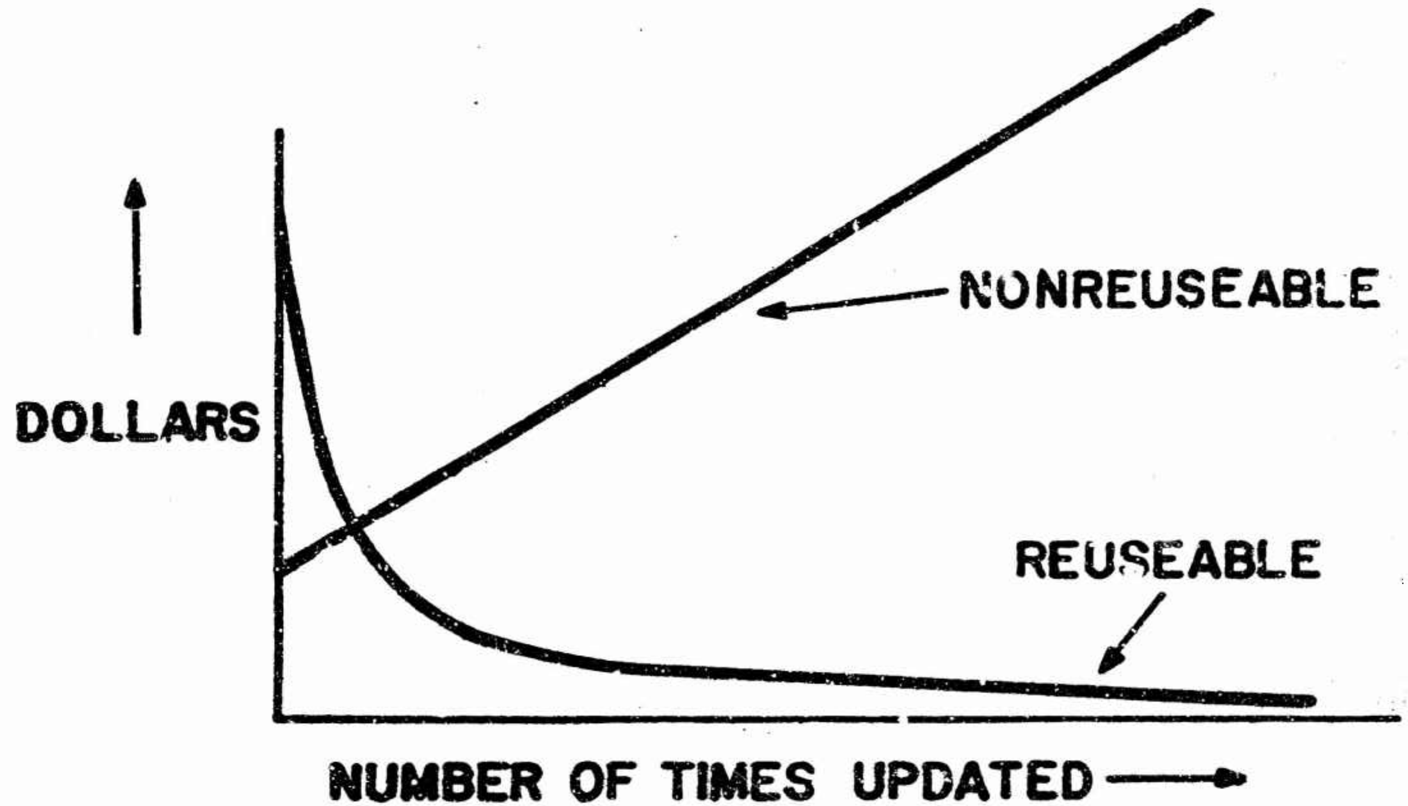


FIGURE 7

shows very clearly and plainly that for lowest cost, reuseability is a prime factor. In fact, to establish a rugged militarized medium that is nonreuseable may be prohibitive to the Government.

Before determining the Final System Design of the new Unit Record concept, it is necessary to determine a few more specific parameters. One of these is the physical unit record medium size. The Human Factor Engineering Section at the Electronics Command Laboratories conducted studies on various sizes of Unit Records to determine the ideal size to meet manual handling requirements. The results of manually sorting and searching various sized Unit Records by length, width, and area are shown in Figure 8, "Manual Handling Search Speed versus Unit Record Size". The results from these tests and studies indicate a Unit Record from a physical handling point of view should be on the order of 3 by 5 inches.

So far it has been determined that the new Unit Record should contain on the order of 1000 machine readable characters, but it has not been determined how many man readable characters it should be capable of recording. Also, no consideration had been given to the ideal size of the Unit Record for machine handling. To look at both of these problems a graph was plotted of the Unit Record physical size versus the Automatic Search Time and the man readable file size. This is shown in Figure 9, "Unit Record Size versus Search Time and File Size". At the upper portion of the graph, the Machine Search Time versus Unit Record Size, it can be seen that the smaller the Unit Record size becomes the longer it takes to search out one Unit Record. Also the larger the Unit Record size becomes, it is more difficult to find a particular document because of its gross dimensions. But for a certain range of Unit Record sizes the machine search time is relatively constant. The lower portion of the graph shows the relationship of the number of man readable characters to the physical size of the Unit Record. This is a linear relationship starting at a certain area size allowing for borders and space between characters. Also taken into account is space allocated for the machine readable characters, an image, and a signature. The important thing to observe here is the trade-offs between these two graphs. At this point it is necessary to determine the number of man readable characters available. Inter-relating these curves, it can be seen that a Unit Record size consistent with that size for optimum manual handling specifies 150 characters and from the graph allows a small period of time for Machine Searching. The dotted line indicates the approach taken. It is interesting to note that the size for manual handling chosen earlier is also very consistent with a unit record size for fast handling by a machine. This in turn allows a reasonable number of man readable characters to be specified, 150 characters.

III RESULTS

From this Operational Research Analysis it was possible to write a specification on a new Unit Record Technique to be used as the requirements for an exploratory development program on a new Unit Record concept. This specification includes new Unit Record Characteristics such as the four functional areas and their capacity, see Figure 10, "Abstract Representation

MANUAL HANDLING SEARCH SPEED VERSUS UNIT RECORD SIZE

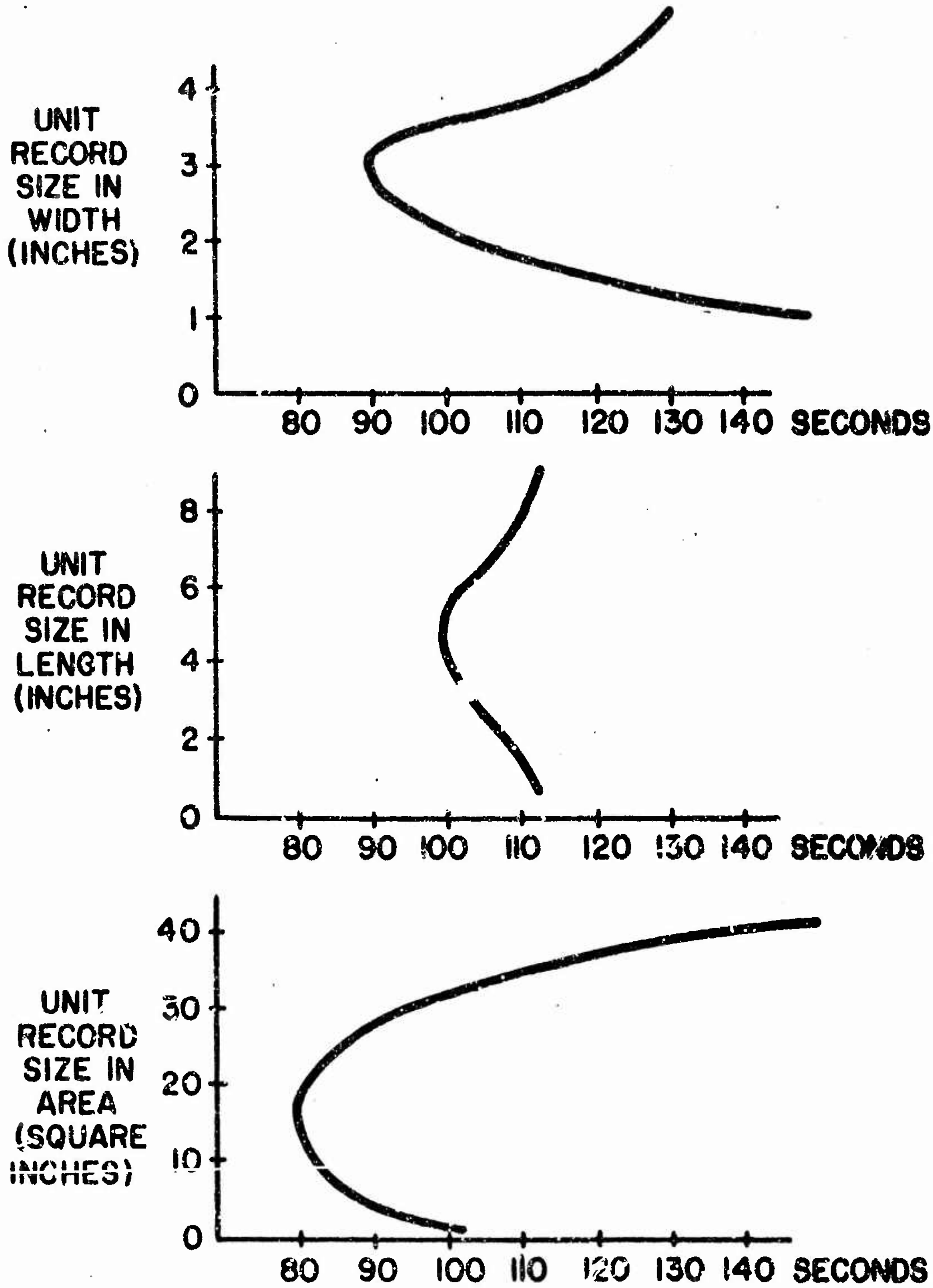


FIGURE 8

UNIT RECORD SIZE VERSUS SEARCH TIME AND FILE SIZE

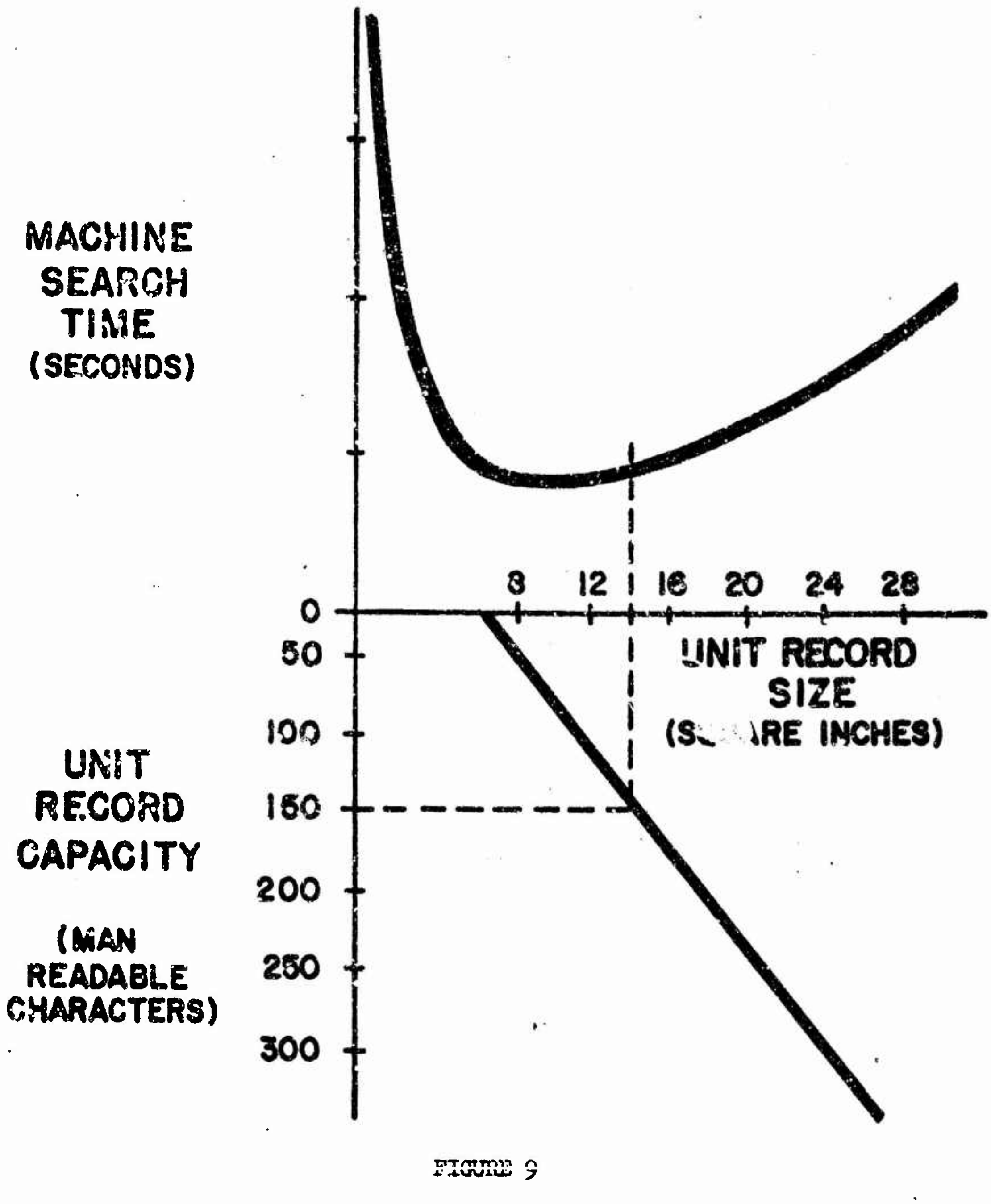


FIGURE 9

of Unit Record". The machine readable and man readable capacity in characters was directly determined from the Operations Research analysis made. The other sectors or areas came from the user requirement studies. Another major factor established for the Unit Record specification is the size of the Unit Record medium. As shown earlier, it is on the order of 15 square inches with a form factor of 5 to 3, that is length to width. This ratio is 1.67. Another determination was the new Unit Record concept should have emphasis of low cost. In fact, high priority should be placed on a low cost reuseable unit record approach.

At present, from this specification two external contracts have been let on Unit Record techniques. There is also an in-house study being conducted on certain Unit Record mediums. It is felt at the end of these one year contracts and internal study, it will be possible to see certain advantages in one particular technique and may offer the first step toward implementing Unit Records. This points the way toward a new technologically updated Unit Record for the Army.

Some possible uses for Unit Records besides the normal use for input/output data and programming would be for dog tags, pay checks, and all hard copy documents. A new Unit Record technique could conceivably revolutionize the entire ADP community.

ABSTRACT REPRESENTATION OF UNIT RECORD

MACHINE READABLE AREA MINIMUM OF 1000 CHARACTERS	MAN READABLE AREA MINIMUM OF 150 CHARACTERS
IMAGE AREA 1-1/2" by 2"	AUTHENTICATION AREA 1-1/2" by 2-1/2"

FIGURE 10

BIBLIOGRAPHY

Bourne, Charles P. "Methods of Information Handling" 1963

Casey, Robert S. "Punched Cards, Their Applications to Science and Industry" 1958

Kent, Allen "Textbook on Mechanized Information Retrieval" 1962

VanNess, Robert G. "Principles of Punched Card Data Processing" 1962

Williams, William F. "Principles of Automated Information Retrieval" 1965

MILITARY DOCUMENTS

"Ammunition Supply as Input to CCIS-70" CDC January 1964

"An Automatic Data Processing Input-Output Equipment Study" Melpar July 1959

"Automated Data Base Inventory" Dept of Army 1 May 1964

"Civilian Personnel Management Information System" U. S. Army Missile Support Command November 1964

"Development of a Military Pay System for the Army in the Field 1965-1970" CDC November 1964

"Development of Operational Systems Requirement for Military Police Functions as Input to CCIS" CDC April 1965

"Legal Implications of Projected Automation of Personnel and Administration and Logistics Operations in Support of the Army in the Field, 1970" Judge Advocate Agency April 1965

"Logistic System for a Field Army" United Research Services July 1962

"Mildata Studies" Sylvania, Honeywell, NCR July 1964

"NAPALM Documentation for RDT&E" ECOM December 1964

"Personnel Management Automatic Data Processing System for the Army in the Field, 1970" CDC September 1964

"An Ultraviolet Approach to an Integrated Unit Record" CADPL Report No. 12 of the Comm/ADP Laboratory, ECOM, Curry, D., Sumrall, G. 4 January 1966

NEW METHODS FOR PREDICTING ELECTRONIC RELIABILITY

By
Eli J. Dworkin
Staff Engineer
U. S. Army Electronics Command
National Maintenance Point
Fort Monmouth, New Jersey

A paper discussing methods of predicting electronic system reliability without relying exclusively on part failure rate information.

The opinions expressed in this paper are those of the author and not necessarily those of the United States Army Electronics Command.

Abstract

This paper deals with the subject of predicting electronic system reliability. "Current Methods For Predicting Electronic System Reliability" are discussed first to provide the necessary background for the later "Review of New Methods for Predicting Electronic System Reliability". The new techniques reported on are those developed over the last three years which do not depend exclusively on component part failure rates but employ such techniques as:

1. mathematical modeling and use of part parameter distributions to predict failures;
2. "Ball-park" reliability prediction by general categories such as number of active elements, number of cathode ray tubes, etc.
3. reliability prediction by specific parameters such as sensitivity, bandwidth, noise figure, etc., using a general prediction equation for all types of electronic equipments;
4. reliability prediction by use of specific indices for a particular type of equipment. For example, these indices for radios may be tuning frequency range, power output, sensitivity, etc. For radars the indices may be peak power out, pulse repetition rate, etc.

The methods of developing these techniques include circuit analysis, linear regression, Monte Carlo, multiple regression, and non-linear curve fitting.

In addition to summarizing and comparing the major works already published in this area, the author has developed a technique to predict reliability using a method which compares many of the specific indices discussed earlier for a new design with those of an existing design for which the reliability is known. This proposed technique offers the advantage of being able to predict the effect a design change will have on the reliability of a system without recomputing the entire regression equation. In addition, predictions can be made when only a few indices are available and no concern need be given to obtaining specific combinations of indices as was true with the previous techniques.

The functions which express the relationship between the indices and reliability for the proposed technique were obtained by plotting available data on existing Army, Air Force, and Navy equipments.

CHAPTER I

INTRODUCTION

In order for any electronic system development program to be complete, it must include adequate reliability considerations.

One of the necessary steps in a reliability plan is the prediction of the reliability at various stages in the development.

This paper will deal with several methods for predicting electronic system reliability. These methods will include those currently used, new methods recently developed, and an original method developed by the author.

In addition to the reviews of the prediction techniques deficiencies, and outstanding aspects of the prediction methods will be discussed.

For those readers who are convinced of the value of reliability predictions, this paper should prove to be a valuable compendium of the most recent works in reliability predictions.

It is hoped, for those who are undecided as to the validity of reliability predictions, this document will help to further emphasize the importance and utility of these new methods.

It should be noted that the prediction techniques discussed in this paper are constantly being revised and improved.

As a result, the regression equations presented may not be the most current available. If one desires to employ any of these techniques, he should refer to the most recent issue of the paper reviewed to obtain the newest available equation.

CHAPTER II

CURRENT METHODS FOR PREDICTING ELECTRONIC SYSTEM RELIABILITY

A. General

There are many methods currently employed for predicting the reliability of electronic systems. Several of these methods will be reviewed in the following paragraphs. It is essential to understand that reliability predictions, as their name infers, are only estimate of the inherent reliability of a device. This means that no confidence can be associated with the predicted value until adequate testing has been completed.

This fact, however, does not destroy the usefulness of early reliability predictions as a tool to assist the design engineer, as well as offering management an opportunity to evaluate the progress of a reliability program during the development of a new system or equipment.

B. Predictions Based Upon Typical Equipment Classification

This method provides the most elementary type of prediction.2/ In this technique, the reliability of an equipment is determined by a simple comparison to a similar device, performing the same function, for which the reliability is already known. This method might also be utilized to compare and evaluate systems employing similar equipments. From such a study a reasonable picture of how to apportion reliability requirements in the new system, on the basis of existing similar equipments, can be obtained.

C. Predictions Using Active Element Groups (AEG) Count

This procedure bases its prediction at the active element group (AEG) level, where an AEG is considered to be a tube or transistor (active element) and its associated circuitry[9]. The reliability of the equipment is based upon a count of the AEG's, with some average failure rate taken for each AEG. Such an approach is useful to determine the overall effect of adding levels of complexity to an equipment or evaluating some other large change in design, (such as adding an extra amplifier stage) without having to make an exact parts count. The use of AEG failure rates generally assumes some average failure rate for some typical configuration (average complexity, design, and environmental factors). Predictions made at this level will give only an average estimate of reliability and will not take into account particular design details and variations.

D. Predictions by Part Count and Nominal Failure Rate Assignment

This method goes one step farther beyond the AEG level, with the reliability prediction being based on actual part counts[9]. In this case, the actual complexity of the system is taken into account with the prediction being made on the basis of nominal failure rates for various classes of component parts, transistors, tubes, resistors, capacitors, etc. At this level, one is not completely bound to a single failure rate number for a particular component, for often different failure rates are given for various types of a single component (film, composition, or wire wound resistors), and these nominal failure rates can be multiplied by various "K" factors to distinguish between general classes of application (ground, airborne, or satellite).

E. Prediction by Part Count and Assignment of Failure Rates Dependent on Stress Levels

This method offers a greater degree of sophistication in making reliability predictions than the previous three methods[9]. In this case, data are utilized to take into account a number of application and environmental factors for parts. In this manner, reliability data can be used which most closely fit the part application, and environment in making a prediction. In general, data are given in terms of an operating level and an environmental level. Failure rates for various resistors are given in terms of power level and ambient temperature, and for capacitors, failure rates are given in terms of operating voltage and temperature. With this data, a reliability analysis can take into account not only details of basic design such as part types, part counts, and part configuration, but can also be the measure of effect for at least some application and environmental stresses and consider the effect of derating, cooling, etc. This last approach, however, does not end the reliability analysis problem and any prediction made using this type of data from the literature must have its qualifications. First, only the major, but by no means all the part stress factors, are taken into account, and in many cases extrapolations made in the data can be open to question. In addition, variations in part design or quality cannot be explicitly evaluated. In an effort to take more of these factors into account in making a reliability prediction, work is being carried out in the area of failure mechanisms. Here, fundamental physical laws and the study of processes controlled by these laws are used to determine the performance of a component,

but until these advanced techniques are made generally applicable, they cannot be relied upon.

F. Other Miscellaneous Prediction Techniques

In addition to the more common approaches outlined above, other techniques and variations might be mentioned[9,20]. One method utilizes a sampling plan, where in the reliability of a certain portion or number of modules of a piece of equipment is determined. The reliability of the entire equipment is based upon an extrapolation of the sample reliability. This method would appear to be most effective if most modules in an equipment were of a similar nature and complexity, such as logic elements in a computer.

Another technique currently being employed is a variation of the predictions by part count and assignment of failure rates dependent on stress levels. Before the summing of failure rates is performed, it is essential to first construct a functional diagram of the system called a reliability block diagram or a reliability model. The failure rate of each block is then determined by summing the failure rates of the parts. The reliability of each block is then calculated using the expression:

$$R = e^{-t\lambda} = e^{-t/m} \qquad \text{(II-1)}$$

The system reliability can now be calculated using the standard reliability relationships for elements in functional series:

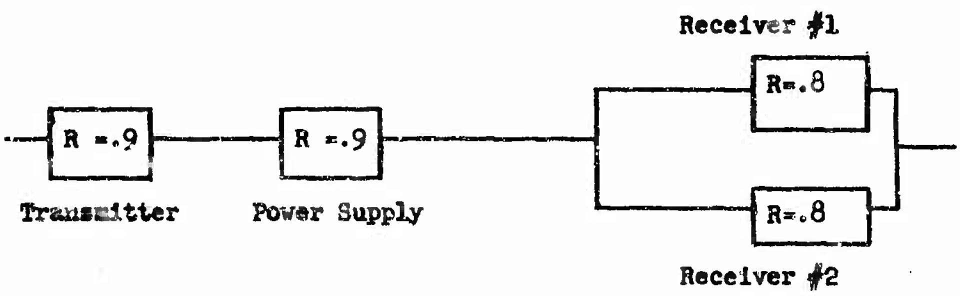
$$R_S = R_1 \times R_2 \times R_3 \qquad \text{(II-2)}$$

and for elements operating in parallel redundancy:

$$R_S = 1-(1-R_a)\,(1-R_b) \qquad \text{(II-3)}$$

example I-1

Find the Reliability of the Functional Block Diagram



$$R = (.9)\,(.9)\,[1 - (1 - .8)\,(1 - .8)]$$
$$R = .81\,[1 - (.2)\,(.2)] = .81\,(1 - .04)$$
$$R = .81\,(.96) = .78$$

G. Analysis of Current Methods

Very little time will be spent analyzing the deficiencies of the first two methods of predicting reliability that were presented in the previous chapter of this paper. It should be apparent that the similar equipment study and the active element group count are only gross estimate techniques, designed to yield a very rough estimate of reliability. These techniques are generally employed during the conceptual stage of a system development when no other technique can be employed because of the obvious lack of sufficient design information. These estimates are highly inaccurate and fail to consider any environmental levels, design configuration, or equipment type (in the case of active element group count).

The techniques which utilize part failure rate information also have their shortcomings, as discussed earlier. The basic fault with these techniques lies in the fact that the predictions made are only as good as the failure rate data used. Some of the more common failure rate sources are:

1. MIL-HDBK-217, 8 August 1962
2. RADC Reliability Notebook, 31 December 1961
3. Failure Rates by D. R. Earles and M. F. Eddins
4. WADD Technical Report 60-330 Supplement I - November 1961
5. Electronics and Electrical Reliability Handbook, Martin Baltimore
6. FARADA Data
7. Battelle Data Bank
8. Guide Lines for Reliability Prediction of Electronic Equipment RDP224, Collins Radio Company

Although many electronic parts are listed in the above sources, there are still many parts for which no accurate failure rate data has been assembled. These parts include:

1. Low-density components for which adequate operating time has not been accumulated to accurately determine a generic failure rate.
2. New parts which are being developed.
3. Parts having an extremely low failure rate so that unreasonable test time would be required to verify this failure rate. Attempts have been made to overcome this lack of data by subjecting parts to accelerated testing conditions and based on these results extrapolate the data to predict what the expected failure rate will be in the normal operating environment. These accelerated testing results have met with many serious criticisms. These include such arguments as the data used does not clearly define a single linear trend, so that extrapolation in the normal operating areas is not accurate. Arguments have also been extended that the mode of failure induced by the accelerated test levels are different than those actually encountered in use. As a result, any "K factor" derived from accelerated testing is not realistic.

There is considerable difficulty in relating "random" failures to a particular application. The so-called random failures usually occurring in equipment are actually caused by certain components being subjected to intolerable transient overstresses arriving at the components at infrequent intervals of time. We can think of this process in terms of energy being transferred into the component

from its total environment at a rate which exceeds the components ability to dissipate it and building up energy concentration within the component beyond its ability to withstand. Examples are voltage punch through in diodes, over-heating in semi-conductor junctions, vibration fractures in welded connections, "hammers" in hydraulic lines and dielectric breakdown in electric lines.

It must also be realized that in most cases the published failure rate data is supplied by the manufacturer of the part under consideration. Since it is reasonable to assume that all manufacturers would like their parts to be accepted as the best or most reliable, they may have tendencies to "pad" their test results or discard test data which gives an unfavorable appearance. The failure rate data which is finally published is generally an average of many manufacturers data over a span of several years. As a result, this average generic failure rate may actually not be indicative of any one part which is being employed in a particular design undergoing evaluation.

From this discussion, it becomes evident that some technique is required to overcome the deficiencies of the current methods or to compensate for their inadequacies.

Reliability prediction, as a tool applicable to the first stages of system design, has changed in character from early days of piece-part stress analysis. It is now necessary to use specific and tailored approaches particularly when dealing with modern complexes such as weapon control and navigational systems. Also, improvement in electronic piece-part quality of construction during the last five years has made conventional stress analyses more routine and restricted to the indication of pure catastrophic tendencies at the piece-part level. Simultaneously, equipment complexity is compounded by computer oriented control and measurement devices making even more necessary the sophistication of system planning and analysis.

Thus we see two significant factors emerging in modern reliability:

1. The necessity of using a system approach since the majority of equipments eventually become integrated within some electronic complex.

2. The importance of functionally oriented reliability analysis during early hardware design stages and subsequently throughout all other phases of a project.

It is believed that the following studies have encompassed both these factors and that the techniques developed will provide a means of relating the reliability of proposed equipment and system design concepts to that of alternate choice or choices. In this way, knowledgeable quantitative trade-off studies can be conducted and decisions reached that will ensure that proposed systems meet their required goals.

CHAPTER III

REVIEW OF NEW METHODS FOR PREDICTING ELECTRONIC SYSTEM RELIABILITY

A. General

This chapter of the paper will be devoted to a review of some of the most recent techniques for predicting electronic system reliability. The methods to be discussed will be:

1. Design Stage Reliability Prediction for Electronic Equipment.
Prepared by J.B. Tommerdahl and A.C. Nelson of Research Triangle Institute, Durham, North Carolina, October 1963.
2. System Reliability Prediction by Function.
Prepared by Arinc Research Corp, Washington, D.C., August 1963.
3. System Reliability Prediction by Function.
Prepared by Federal Electric Corporation, Paramus, New Jersey, May 1963.
4. Program to Establish Review-Point Criteria for Reliability Monitoring.
Prepared by Hughes Aircraft Company, Fullerton, California, July 1964.

B. Design Stage Reliability Prediction for Electronic Equipment

The basic objective of the Research Triangle Institute study was to provide guidelines for a technique which can be utilized to realistically assess the reliability of electronic circuits in the early stages of equipment design[20]. An experimental program which utilized two types of transistorized circuits was conducted to demonstrate the feasibility of the technique.

Basic Concepts. In predicting the reliability of electronic circuits, it is necessary to consider performance failures as well as catastrophic failures. A very pronounced change in a performance parameter value within a relatively short period of time is generally termed a catastrophic failure. The mode attributable to performance failure is the result of relatively slowly changing component parameter(s). Catastrophic failures are random in nature and are attributable, in general, to faults in the basic materiels and manufacturing processes and to environmental/electrical over-stress.

In order to realistically predict performance failure of a circuit, it is necessary to be able to predict performance parameters as a function of the pertinent variables. These variables, which are the component parameters in this case, exhibit variations around nominal or mean values and may be described by probability distributions. Such distributions are usually time dependent; i.e. the variances and means will shift with time due to normal physical and chemical processes. Thus one must utilize performance prediction methods which will enable one to estimate the time dependent performance parameter distribution.

Performance Prediction Methods. If one selects an appropriate performance criterion for the circuit and relates performance to the component parameters, then an estimate of the probability of performance failure due to degradation or drift in the circuit parameters can be obtained. This is the aspect of reliability estimation with which the Research Triangle document is primarily concerned. Estimation of the probability of performance failure can be performed using one or more of the following methods of analysis:

1. Propagation of error by means of Taylor Series approximations.
2. Computer simulation such as Monte Carlo using known transfer functions.
3. Empirical mathematical modeling;
   a. Random selection of parameters;
   b. Statistically designed selection of parameters.

If the transfer function relating the performance of the circuit to the component parameters is available, either method (a) or (b) may be used.

Method (a) yields an estimate of the mean and variance of the performance by means of a Taylor Series expansion of the transfer function. By this procedure the performance is approximated by a linear function using only the first order terms in the expansion. One should check the adequacy of this linear approximation for the range of variation of the component parameters which one expects due to manufacturing variation and variation due to aging or degradation. In some cases, second degree and higher order terms may be required.

Method (b) yields a direct estimate of the distribution of the performance by computer simulation utilizing known transfer functions. In this case, various distributions of the component parameters are assumed, random numbers with the assumed distributions such as normal, logarithmic normal, gamma, chi-square are generated, and then by computation of the performance from many such values an estimate of the performance distribution is obtained. The precision of the estimated performance distribution can be expressed as a function of the sample size, i.e. the number of the Monte Carlo or simulation trials.

In a circuit composed of discrete component parameters, it is often possible to describe mathematically the performance of the circuit in terms of the component parameters. Such a mathematical expression is not usually adequate, even for a simple circuit, since the contribution of each parameter to the performance function is not generally obvious or readily ascertainable. In cases such as those just cited or where the theoretical derivation of the transfer function is impractical, the approach is to develop an empirical mathematical model or empirical transfer function.

There are two distinct empirical procedures one can use to estimate the performance distribution such as the mean and variance. The first one is to select the components at random, build the circuits and measure the performance of each circuit under certain conditions. Then the observed distribution of the performance measurements serves as an estimate of the actual performance distribution. Alternatively, an empirical model may be derived from the data, and then an estimate made of the performance distribution. The second method is to measure the parameters of a representative sample of each component, then build the circuits using specific components selected according to a statistically designed experiment; next, obtain an empirical model and then estimate the performance distribution. The second method is the one described in the RTI report. The estimate given by the second method is limited only by the ranges of the parameters used and hence is applicable to most general distributions. Furthermore, assuming that one can obtain an empirical model, the second procedure will yield more precise performance prediction as a result of the experimental design.

C. System Reliability Prediction by Function (ARINC)

The basic objective of the Arinc Research Corporation study was to provide a technique for predicting the reliability of electronic systems and equipment during the pre-design stage before detailed design information is available[18].

The prediction procedure is based on field reliability data of Air Force ground electronic systems, supplemented by data from shipboard electronic systems and, in one case, by the results of a reliability prediction conducted according to procedures given in MIL-HDBK-217. In all, reliability data on 51 systems were analyzed to derive the numerical relationships used in the methods outlined in the Arinc report.

Multiple regression analysis was used to determine the simultaneous influences of the many system parameters. As a part of the analysis, the statistical significance of the contribution of each parameter on reliability was determined.

The system parameters considered in the Arinc study are presented in Table 1 followed by a review of the Prediction Technique.

The results of the investigation led to a prediction procedure based on observed field conditions. The inclusion of a confidence interval as an integral part of the predictions insures that the resultant predictions realistically represent the wide variations to be met in the field.

Summary of Procedure. The prediction procedure comprises six major steps:

1. Selection of appropriate prediction equation based on the amount and type of information available;
2. Numerical evaluation of system parameters;
3. Solution of prediction equation for $\ln \hat{\theta}$, where $\hat{\theta}$ = predicted Mean-Time-Between-Failures (MTBF);
4. Transformation from $\ln \hat{\theta}$ to $\hat{\theta}$ ($\hat{\theta} = e^{\ln \hat{\theta}}$);
5. Computation of confidence interval in terms of $\ln \hat{\theta}$ (confidence interval = $\ln \hat{\theta} \pm K$);
6. Translation into terms of θ (confidence interval is $\hat{\theta}/g$ to $g\hat{\theta}$, where $g = e^{K}$).

After the completion of Step 6, the effects of redundancy may be considered, if desired. However, detailed steps for evaluating the influence of redundancy were not included, since redundancy may be employed in so many diverse forms that it was not possible to take all possibilities into account as part of the procedure.

D. System Reliability Prediction by Function (Federal Electric)

The purpose of this study was the development of electronic system oriented prediction techniques by correlating significant functional characteristics to operational Mean-Time-Between-Failures (MTBF's) so that system reliability can be planned and allocated during the early design stages[19].

It was intended that these techniques be used by system design engineers to estimate the reliability of a system during the early phases of the design cycle when no firm data is available concerning the ultimate configuration to be employed on the total part compliment of the system. The techniques provide a means of relating the expected reliability of one or more proposed design concepts to that of an alternate choice or choices so that knowledgeable quantitative trade-off studies can be conducted and decisions reached that will assist in determining if the proposed system design meets the required goals.

TABLE 1

DEFINITIONS OF SYSTEM PARAMETERS

Parameter Designation (Ranges of Validity)	Definitions
$X_1 = \ln C - 2.303$ $C = 1.0 \times$ No. of electron tubes (all types) $+0.3 \times$ No. of analog transistors $+0.4 \times$ No. of solid state power rectifiers $+0.006 \times$ No. of solid state digital diodes $+0.02 \times$ No. of digital transistors $+0.1 \times$ No. of solid state analog diodes $+10^{-5} \times$ No. of magnetic core devices (e.g. magnetic amplifiers, memory cores, etc.; passive devices, such as transformers, are not included). Upper limit = 8.087 (C=32,488) Lower limit = 1.609 (C=2)	Natural logarithm of system complexity (C) minus 2.303. Complexity is measured in terms of an adjusted active element count. A worksheet is provided in Section 2.4 to aid in the numerical evaluation of X_1. The number of each type of active element in the systems of this study ranged from zero to the following maxima: Tubes -- 3,037 Analog transistors -- 4,167 Solid state power rectifiers -- 204 Digital diodes -- 25,000 Digital transistors -- 3,083 Analog diodes -- 692 Magnetic cores -- 16,200 An element is considered as digital in application if the function depends primarily on the existence of a pulse, rather than on the shape of a signal.
X_2 Upper limit = 11 Lower limit = 0	Number of cathode-ray tubes in the system.
X_3 Upper limit = 2 Lower limit = 0	Number of transmitters in the system.

(continued)

TABLE 1 (continued)

DEFINITIONS OF EQUIPMENT PARAMETERS

Parameter Designation (Ranges of Validity)	Definitions
X_4 = Ln F - 2.303 Upper limit = 6.908 (F = 10,000Mc) Lower limit = 9.211 (F = 0.001Mc) F = Frequency (megacycles)	Natural logarithm of highest frequency characteristic of the system, in megacycles per second, minus 2.303. (In many cases, the highest frequency characteristically found in a system must be estimated on the basis of engineering judgment, as in the case of a radar repeater. The influence of this parameter on the result is such that an error in the estimate of frequency of a factor of two or three will be acceptable.) In the case of tunable equipment, the upper limit of the tuning range will be the estimated highest characteristic frequency; e.g. in the case of a 2[illegible]/400 Mc radio set, the frequency to be used is 400 Mc. The units of frequency in the equation are megacycles.
X_5 Upper limit = 2 Lower limit = 1	The value of X_5 is 1 for systems which are primarily digital in nature, and 2 for systems which are primarily analog in nature. In the case of mixed equipments, the choice is made according to the nature of the majority of the important functions. The distinction between digital and analog is the same as noted in the definition of X_1.
X_6 Upper limit = 1 Upper limit = 0	The value of X_6 is 1 if the system has a steerable antenna. Otherwise, the value is zero.
X_7 Upper limit = 5.489 (242 kV) Lower limit = 4.200 (0.015 kV)	Natural logarithm of the maximum dc voltage in the system, in kilovolts.

(continued)

TABLE 1 (continued)

DEFINITIONS OF EQUIPMENT PARAMETERS

Parameter Designation (Ranges of Validity)	Definitions
X_8 Upper limit = 5.989 (399 kW) Lower limit = 3.412 (3.033 kW)	Natural logarithm of total rated power consumption, in kilowatts.
X_9 Upper limit = 0.984 Lower limit = 0	X_9 is the ratio of the tuning range of the system to the highest characteristic frequency: $X_9 = \frac{\text{Tuning range (Mc)}}{\text{Frequency (Mc)}}$ For equipments with no tuning provision, X_9 has the value of zero.

These correlation studies were applied to broad system areas and are applicable to different equipment functional levels, ranging as specified from command to data acquisition, to receivers, and transmitters. The procedures are not intended to replace the conventional methods of reliability prediction such as stress analysis as is contained in the RADC Reliability Notebook TR-58-111. Instead, they are to be used to augment them in the early pre-design planning stages at a time when specific part selection has not been defined.

Correlation studies in the program were categorized under:

1. Radar
2. Ground Communication
3. Electronic Data Processing

In the first two of the three basic categories, namely Radar and Communication, a number of applicable systems were selected for functional study and correlation. Selection of specific systems was made on the basis of available field operational and failure data, on the general range of the individual systems performance and character, plus consideration of current and standard design practice and technical development in the state of the art.

E. Program to Establish Review - Point Criteria for Reliability Monitoring

The objective of the Hughes Aircraft Company study was to develop a reliability estimation technique which will utilize reliability indicies of equipment characteristics other than component part populations[12]. These indices were developed by a statistical analysis. The prediction equations and constant multipliers were obtained by a linear multiple regression computer analysis.

The equations that were developed were for six specific types of equipment. These equipments were:

1. CRT Displays
2. Radars
3. Radio Transmitters
4. Radio Receivers
5. Radio Transceivers
6. Buffering Equipment

The characteristics which were chosen as possible indicators were:

1. MIL-HDBK-217 Reliability handbook estimate of MTBF
2. Prime power input
3. Power dissipation per unit volume
4. Peak power output
5. Average power output
6. Efficiency
7. Circuit bandwidth
8. Turning range
9. Sensitivity
10. Ratio of types of assemblies used once to total types of assemblies
11. Ratio of connectors to total parts
12. Packaging density
13. Power per active element
14. Power supply regulation
15. Acceleration voltage
16. Minumum sweep range

17. Noise factor
18. Gain bandwidth
19. Selectivity
20. Ratio of operator adjustments to total adjustments
21. Ratio of moving parts to total parts
22. Ratio of standard parts to total parts
23. Thermal hot spot temperature
24. Clock pulse rate
25. Tolerance of clock pulse rate
26. Storage capacity
27. Bits per word
28. Existence of contractual reliability requirement

CHAPTER IV

COMPARISON OF NEW PREDICTION TECHNIQUES

A. Limitations

The technique developed by Research Triangle Institute is limited in its range of applications because of the complex mathematics required. In order to employ their technique one must design an adequate experiment, build sufficient circuits to obtain measurements, take measurements and then using a computer obtain the desired constants for insertion desired answer. Since this technique is based at the circuit level the process must be repeated many times over in order to analyze an entire equipment or system. If the circuits which are being employed are complex, the number of circuits required would be tremendous.

Since the process is quite time consuming, it does not appear to be workable at the level of the design engineer. If the designer has enough time and information to build actual circuits, he could just as well build an experimental model and run some tests to determine an estimate of the reliability.

The prediction techniques of Federal Electric, Arinc, and Hughes Aircraft Company are limited as a result of the statistical analysis used to determine the regression equations.

In a statistical analysis such as used in the development of these procedures, the validity of the predictions are subject to doubt if the range limits of the input data are exceeded. The assumption is made for these procedures that the design of a new system will comprise elements and principles similar to those of the designs considered in the analysis. Accordingly, the procedures are of questionable value for use with systems whose parameters are outside the ranges encountered in the study.

B. Validity of Procedures

Despite the disadvantages of the Research Triangle Institute technique, their study did make a major contribution in the area of recognition of basic differences between drift failure and catastrophic failure. Until now, the assumption has been made that electronic equipments do not wear out, and that they display a constant hazard probability. As a result of this study, it becomes apparent that electronic components do display a drift characteristic which may result in operation outside a specified limit which would then constitute a failure. If for some reason, one desires to separate degradation failure probability from catastrophic failure probability, this technique is one of a very few which offers a solution.

The other prediction equations, being based on field reliability data, can be expected to give fairly accurate indications of the performance of new systems in the field environment. In these studies, any complaint against the systems which necessitated corrective maintenance action was assumed to be a failure, whether the cause was fundamentally due to the design of the equipment or to the environment, maintenance policy or other external factor. For this reason, it may be anticipated that the MTBF's predicted from the equations may be somewhat lower than MTBF's computed from data based only on catastrophic failures.

The question arises as to the extent that past experience (as represented by field data) can be utilized in the prediction of the behavior of new designs, particularly since the standard deviations and confidence limits are really only measures of how well the regression equation fits the available data. The validity of the prediction procedure described herein, and the statements concerning the confidence intervals, rest on the assumption that the new system has the same distribution of characteristics as those systems which were used to generate the data for this program. An assumed corollary is that the prediction equations are valid for systems designed and fabricated according to the engineering practices of the time interval (1949-1962) associated with the systems of these study programs.

C. Engineering Significance of Coefficients

The development of the prediction equations have been based on a statistical approach, without rigorous analysis of the engineering or cause-and-effect relationships between the system parameters and reliability. While some of the relationships can be justified intuitively, it is not valid to assign engineering importance to the coefficients. (For example, while a correlation has been shown between power consumption and MTBF, a high-power consumption is not necessarily the cause of a lower reliability. It may be that increased power consumption is an additional measure of complexity, which in turn influences reliability.)

For these reasons, it is not valid to attempt to combine the individual prediction equations in an attempt to arrive at new results or to gain insight into the fundamental cause-and-effect relationships affecting equipment reliability. While these considerations impose limits on the interpretation of the results, they do not adversely affect the ability to predict reliability.

D. Comparison of Techniques

The following table will compare the Research Triangle Institute, Hughes Aircraft Company, Arinc Research Corporation, and Federal Electric Company, Division of ITT, reliability prediction techniques. The points of comparison will be:

1. Method of estimating reliability.
2. Applicable stages of development when techniques are to be employed.
3. Type of equipments for which the techniques can be employed.
4. Analysis approach used.
5. Analysis technique used.
6. The reliability index used.
7. Reliability indicators considered in the techniques.

TABLE 2

COMPARISON OF RELIABILITY ESTIMATION TECHNIQUES

Point of Comparison	RTI	Hughes	Arinc	FEC	Proposed Original Techniques
Method of estimations	Mathematical model	Binder weighted function of reliability indices	Linear logarithmatic weighted function of reliability indicators	Nomographs & curves of failure rates	Product of factors obtained from linear graphs
Applicable stages of development for technique	Post-Circuit Design	Pre-Circuit Design	Pre-Circuit Design	Pre-Circuit Design	Pre-circuit Design
Equipment type	All Electronic circuits	Specific types listed	General all types of ground electronic equipment	Specific for radars and ground communication equipment	Specific types listed
Analysis approach	Comparison to known transfer functions	Statistical analysis of existing equipment	Statistical analysis of existing equipment	Statistical analysis of existing equipment	Statistical analysis of existing equipment
Analysis technique	Method of least squares	Multiple regression	Multiple regression	Correlation and nonlinear curve fitting	Multiple regression
Reliability Index	Percent probability that output value lies between acceptable limits	MTBF	MTBF	MTBF	MTBF

(continued)

TABLE 2 (continued)

COMPARISON OF RELIABILITY ESTIMATION TECHNIQUES

Point of Comparison	RTI	Hughes	Arinc	FEC	Proposed Original Techniques
Reliability indicators considered	Part parameter distributions	Handbook predictions Power dissapation per unit volume Prime power input Power supply regulation Minimum sweep range Acceleration voltage Peak power output Average power output Receiver noise figure Receiver gain bandwidth Ratio of types of assemblies used once to total types of assemblies	Number of active elements by type Number of cathode ray tubes Number of transmitters Highest frequency Analog or digital type equipment Steerable or fixed antenna Maximum DC Voltage Rated power Ratio of tuning range to highest frequency	Active element groups Noise safety margin Peak power Carrier-to-noise ratio Number of multiplex Channels	Ratio of handbook prediction old design to handbook prediction of new design Ratio of power dissapation per unit volume of old design to new design Ratio of prime power input old design to new design Ratio of peak power output old design to new design Ratio of average power output old design to new design Ratio of receiver noise figure old design to new design Ratio of receiver gain bandwidth old design to new design

(continued)

TABLE 2 (continued)

COMPARISON OF RELIABILITY ESTIMATION TECHNIQUES

Point of Comparison	RTI	Hughes	Arinc	FEC	Proposed Original Techniques
		Ratio of operator adjustments to total adjustments			Ratio of operator adjustments old design to new design
		Ratio of moving parts to total parts			Ratio of moving parts old design to new design
		Ratio of connectors to total parts			Ratio of connectors old design to new design

CHAPTER V

RELIABILITY PREDICTION BY INDEX COMPARISON

A. General

This chapter of the paper will be devoted to presenting a new method of predicting system reliability which is the original work of the author. This new technique will employ a system of comparing the reliability of a new design to that of an existing design for which the reliability is known. The basis for comparison will be the ratio of indices of the new design to the old design. Each type of equipment will require a different set of indices, graphs, and weighting factors. The technique will be demonstrated in this paper for radio receivers only. Examples of a few of the indices to be used are:

1. $\dfrac{\text{Sensitivity of new design.}}{\text{Sensitivity of a design with known reliability.}}$
2. $\dfrac{\text{Circuit bandwidth of new design.}}{\text{Circuit bandwidth of old design.}}$
3. $\dfrac{\text{Prime power input of new design.}}{\text{Prime power input of old design.}}$

This technique offers several advantages not available in earlier techniques. It allows a designer to investigate the effect a design change will have on the reliability without having to consider other indices as well. No concern need be given to the number of indices available or to the availability of certain combinations of indices.

B. Limitations on Data

Data, of the type required, to design a statistically sound experiment is not readily available. As a result, the data which is used to illustrate this technique is not meant to be mathematically conclusive. It should be noted that the results are based on twelve samples of equipment, very few of which are duplications of the same item. If one desires to employ this technique, he must first obtain a data set which is complete enough to satisfy his statistical needs and then he can continue on with the employment of the technique. This technique in no way relies on the statistical validity of the data; it is strictly an engineering approach which employs certain statistical tools. If the data is improved to suit these statistical tools, the outcome will obviously be a more accurate prediction.

C. Data

The data presented in Table 3 are examples of the type of data required for employment of this technique. They were obtained from existing Army, Navy, and Air Force equipments.

D. Technique

This prediction technique requires that the data for each reliability index of the old design for which the reliability is known, be plotted as the independent

TABLE 3

RELIABILITY ESTIMATION DATA COLLECTED ON RADIO RECEIVERS

	Equipment No.					
	1	2	3	4	5	6
Field Estimate of Mean Time Between Failures, (MTBF Hours)	132	221	876	1060	2504	2251
Reliability Indicators						
MIL-HDBK-217 Prediction of MTBF (Hours)	192	194	910	720	1230	1840
Power Dissipation/unit Volume (Watts/ft^3)	396	396	205	210	59	59
Sensitivity (Decibels)	10	10	28	14	3	3
Prime Power Input (Watts)	330	330	85	72	88	88
Selectivity (Decibels)	60	60	4	7.5	-	-
Gain Bandwidth (Decibels/ Kilocycle)	1000	1000	8400	2800	4.8	4.8
Bandwidth (Kilocycles)	100.0	100.0	300.0	200.0	1.6	1.6
Tuning Range (Megacycles)	175.0	175.0	5.7	14.8	.59	7.75
Power per Active Element(Watts)	-	-	-	-	0.33	0.31
Packaging Density (lbs/ft^3)	62.2	62.2	47.0	48.0	26.7	26.7
Ratio of Moving parts to total parts	0.038	0.038	0.49	0.67	-	-
Ratio of connectors to total parts	0.068	0.068	0.49	0.57	-	-

(continued)

TABLE 3 (Continued)

RELIABILITY ESTIMATION DATA COLLECTED ON RADIO RECEIVERS

	Equipment No.					
	7	8	9	10	11	12
Field Estimate of Mean Time Between Failures (MTBF Hours)	2060	2065	2120	4050	750	740
Reliability Indicators						
MIL-HDBK-217 Prediction of MTBF (Hours)	1840	1150	1070	1510	1575	480
Power Dissipation/Unit Volume (Watts/ft^3)	59	129	-	-	35	35
Sensitivity (Decibels)	3	0.5	6	8	3	10
Prime Power Input (Watts)	88	155	350	110	140	520
Selectivity (Decibels)	-	-	-	-	-	-
Gain Bandwidth (Decibels/Kilocycle)	4.8	2.7	390	720	30	400
Bandwidth (Kilocycles)	1.6	5.4	65.0	90.0	10.0	40.0
Tuning Range(Megacycles)	30.0	28.0	175.0	175.0	175.0	175.0
Power per Active Element(Watts)	0.31	0.40	1.13	0.49	0.67	0.57
Packaging Density (lbs/ft^3)	26.7	30.8	42.5	35.1	15.0	22.0
Ratio of Moving parts to total parts	-	-	-	-	-	-
Ratio of connectors to total parts	-	-	-	-	-	-

variable. A straight line curve is then fitted to the set of points obtained using the method of least squares. This method sums the squares of the vertical deviations of the plotted points from the straight line curve. The line resulting in the lowest absolute value of the sum of these squares is the best fit (more detail on this method can be obtained from any standard statistical text). If the resulting line is horizontal the index plotted is not a good indicator of reliability and should be discarded. Those indices resulting in a straight line, having a finite slope should be used in later computations.

Those indices which appeared to be significant by the linear plot method are then employed in a linear multiple regression analysis like the one employed by Hughes Aircraft Company presented earlier in this paper.

The results of this multiple regression indicates which indices are significant and what the relative weight assigned to each index should be.

The significant indicies are then plotted as earlier with the index as the independent variable and the MTBF as the dependent variable.

The next step requires that the data be normalized. This is accomplished by dividing all MTBF's by the average MTBF. The factors resulting are the ratios of MTBF s. The independent variable is also normalized by dividing each value of the index by the value associated with the average MTBF. This results in a set of ratios of any index over a reference index.

If an estimate of the effect of a design change is to be made, the ratio of the index of the new design to the old design, for which the reliability is known, must be obtained. Using the curve of ratio of indices versus ratio of MTBF's, a ratio of MTBF's can be obtained. This obtained ratio indicates the effect a design change will have on reliability.

E. Illustration of Technique

This illustration will be a step by step procedure one would follow if he desired to estimate what effect various design changes would have on the reliability of a radio receiver.

Step 1. Obtain data on as many of the indices listed in Table 2 under proposed original technique as possible, and any other indices you may feel will be a significant indicator of reliability for as many radio receivers as possible. This data should be arranged in a table like Table 3 used in this example. A minimum number required would be one more than the number of indices to be examined. This is required in order to get meaningful results from the multiple regression analysis.

Step 2. Plot the obtained data as described in the previous section on Technique and fit the best linear curve to the data. This has been done in Figures 1 and 2 for the data in Table 3. Only those indices are plotted for which there is a complete or near complete set of data. By observing these plots only five indices appear to be significant.

Step 3. The data on the five indices should then be run on the multiple regression program presented in Appendix I. The results of the multiple regression analysis indicated that only four were true indicators of reliability. Power dissipation per unit volume was closely correlated with prime power input and did not contribute any significant information. Since the set of data for power dissipation per unit volume was not complete, it was selected as the index to be eliminated. One of the advantages or using multiple regression is that it removes all biased indicators and assigns heavier weight to those that are better indicators. If all

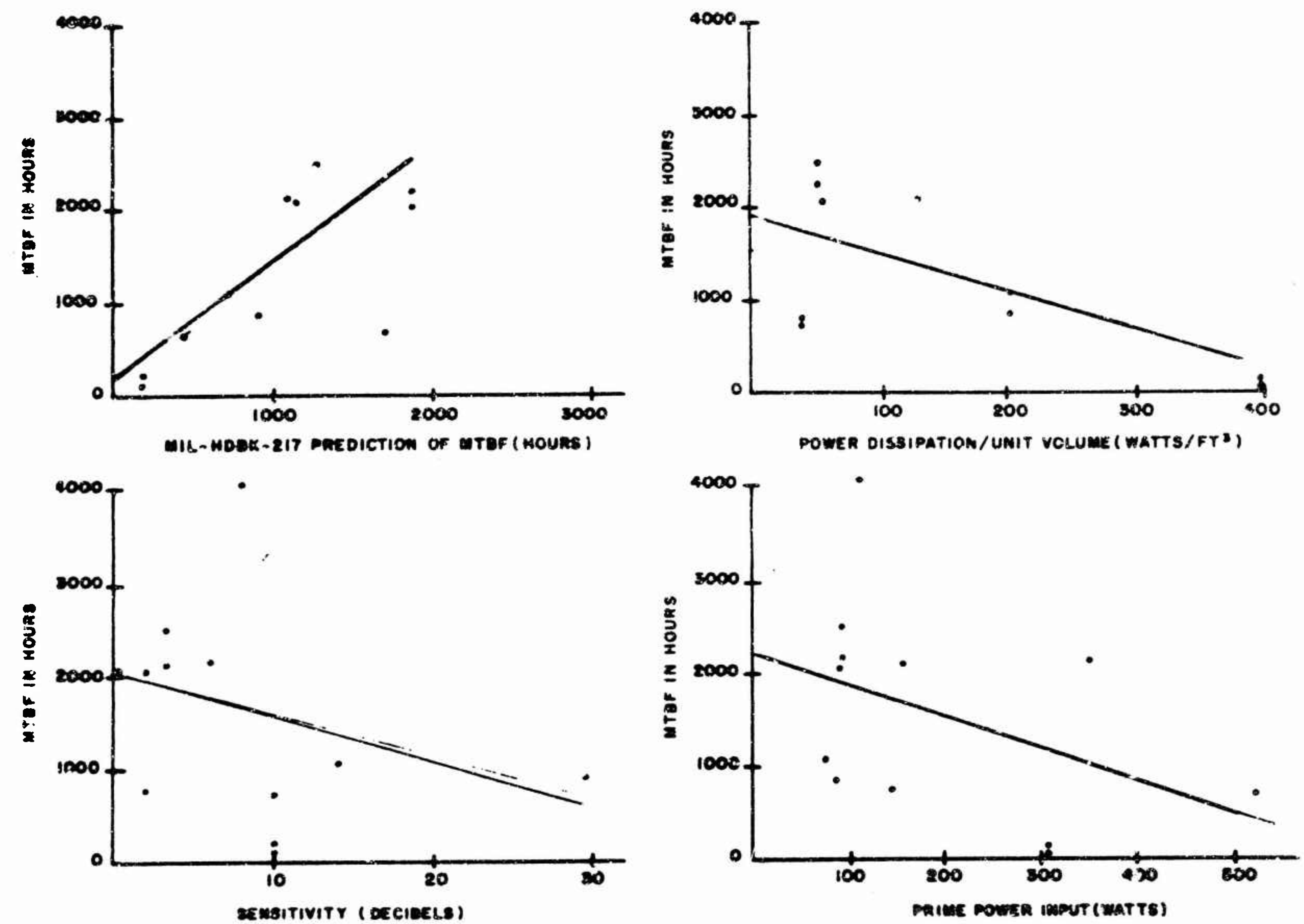


FIGURE I. MTBF VS. RELIABILITY INDICES.

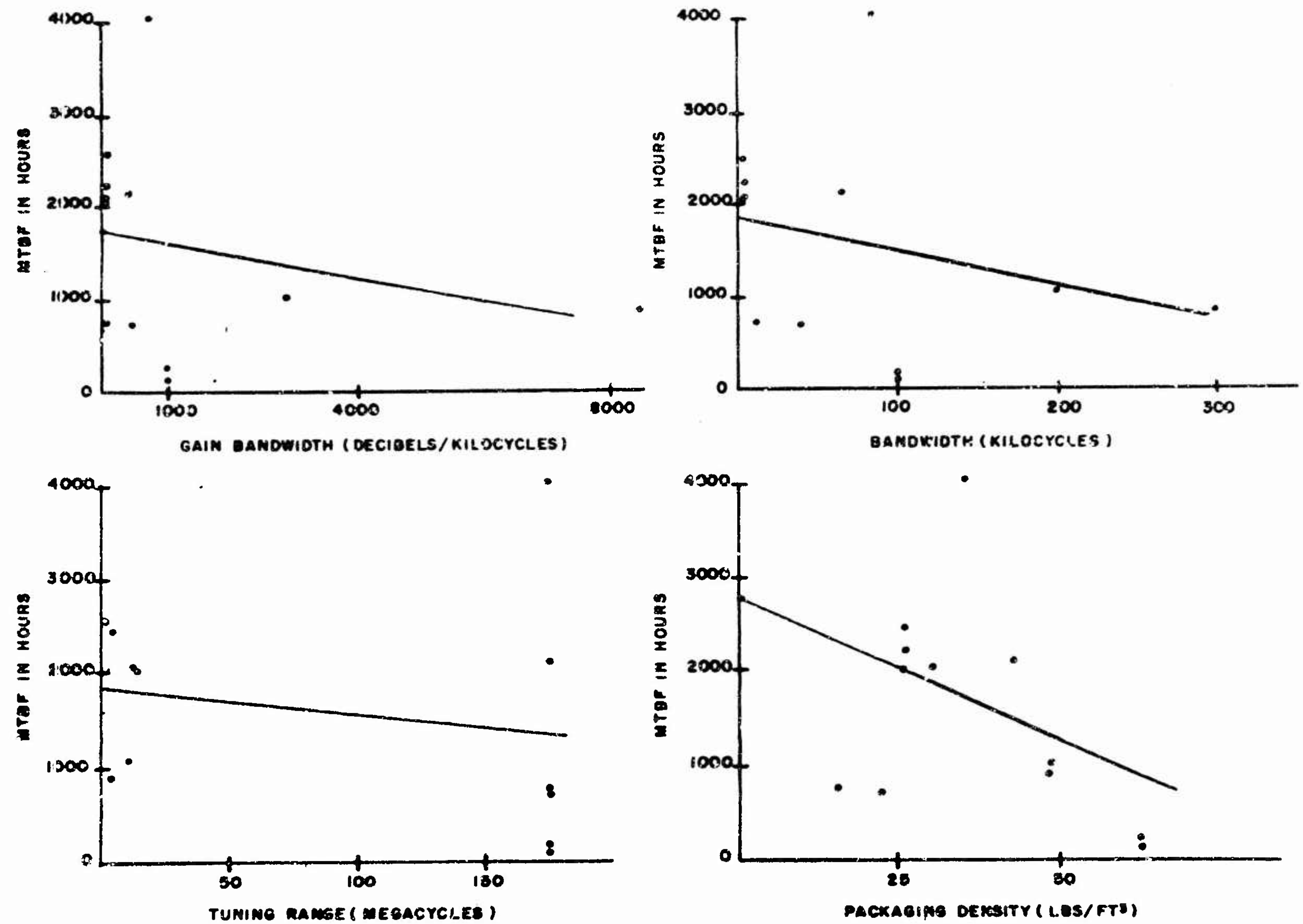


FIGURE 2. MTBF VS. RELIABILITY INDICIES

possible indicators are explored, any false indicators will be eliminated.

Step 4. Replot the true indicators as in Figures 3, 4, 5, and 6 and fit the best linear curve to the data.

Step 5. Normalize the data on both axis as described in the previous section on technique by dividing by the average value of MTBF which is 1,569 hours, and average value of the index being divided and re-number both axis with the new ratio designators as is done in Figures 7, 8, 9 and 10.

Step 6. If a design change is proposed which will vary the prime power input from 200 to 300 watts, examination of Figure 9 indicates that the reliability will be reduced by a factor of .85.

Step 7. If a numerical estimate of the reliability is desired, a known design not used in the data sample should be selected and the ratio of prime power input for the new design and prime power input for the old design should be computed. Enter Figure 9 at this value of index ratio and read the value of MTBF ratio. Multiply this ratio times the known MTBF of the old design and obtain the estimate of the MTBF of the new design.

F. Evaluating Several Indices At One Time

If one desires a more accurate prediction, he can consider the effects of several indices at the same time. The formula used is

$$MTBF = \frac{M_k \sum W_a I_a}{\sum W}$$

where: M_k = MTBF of the known design
W_a = weight of Index
I = ratio of MTBF's for Index obtained from appropriate graph using correct ratio of indices. The weights are obtained from the multiple regression analysis using the technique described in Appendix II. Dividing the result by the sum of all the weighting factors will cancel out the effect of the weighting factor on the magnitude of the final solution.

APPENDIX I

COMPUTER PROGRAM FOR MULTIPLE REGRESSION ANALYSIS*

Identification

Title:	Multiple Regression Analysis
Name:	(MLRG) *
Number:	G2-010M
Source Language:	SCAT
Library Identification:	(MLRG)

*Extracted from bibliography reference 12.

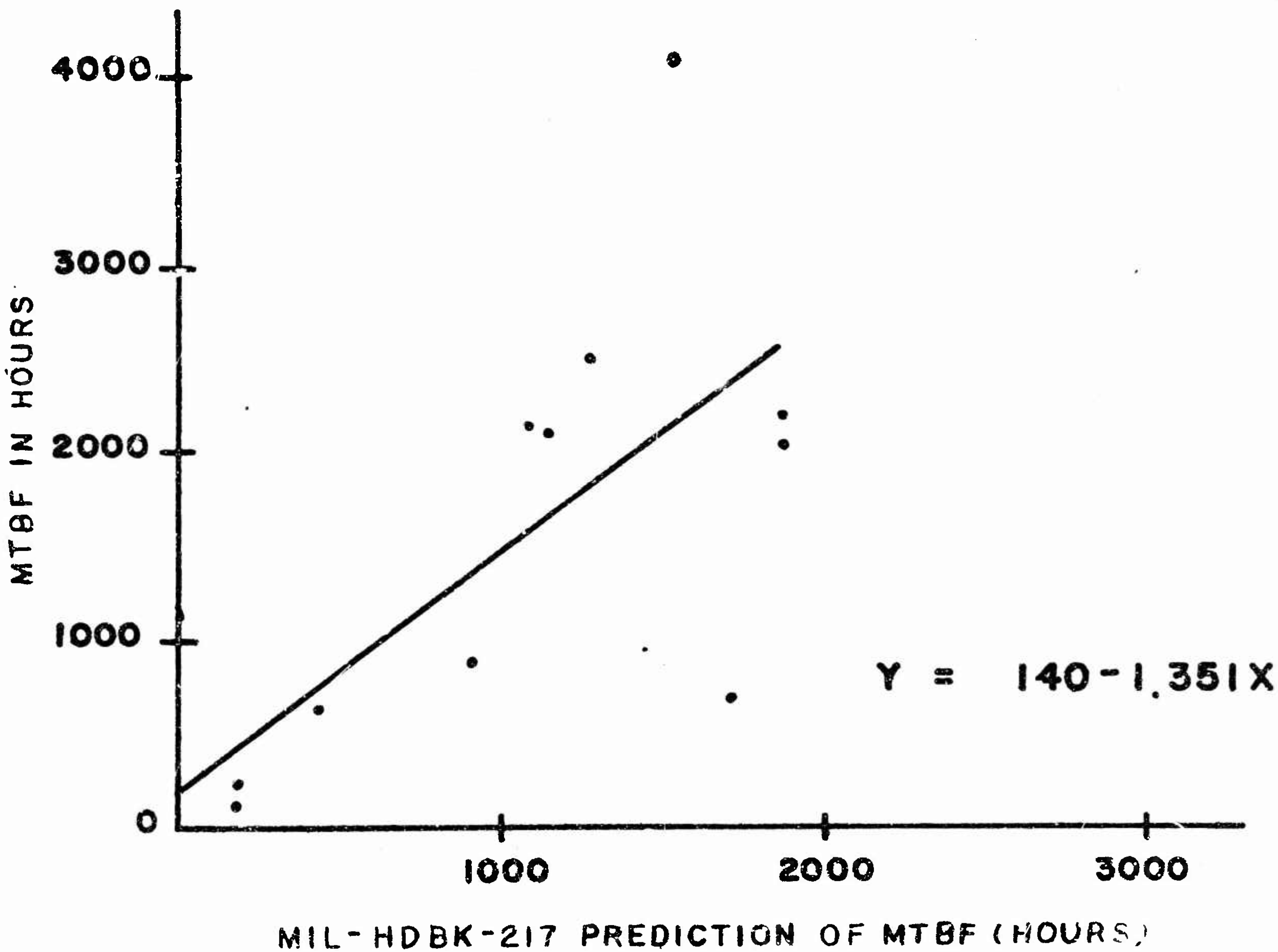


FIGURE 3. MTBF VS. MIL-HDBK-217 PREDICTION OF MTBF.

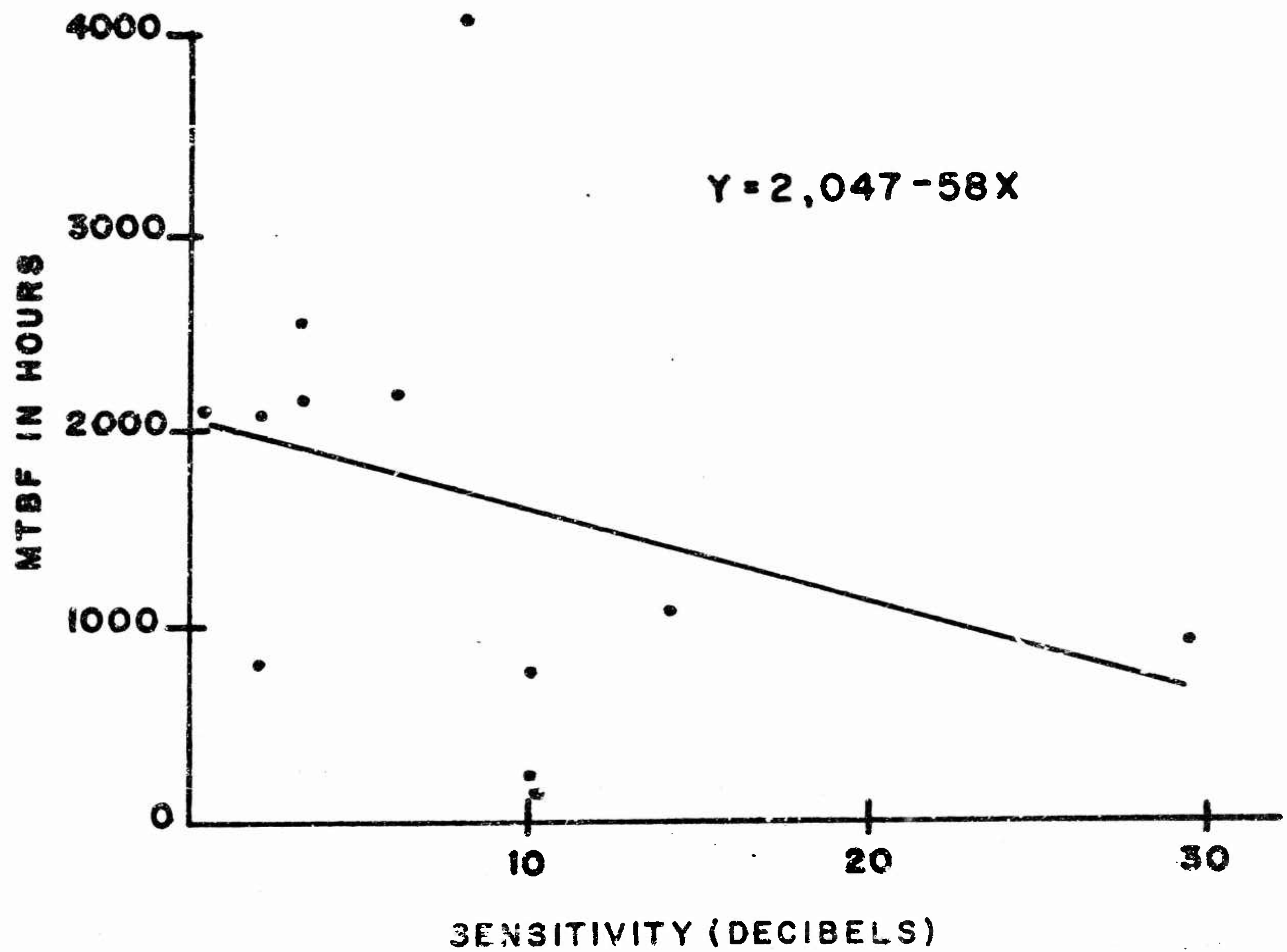


FIGURE 4. MTBF VS. SENSITIVITY.

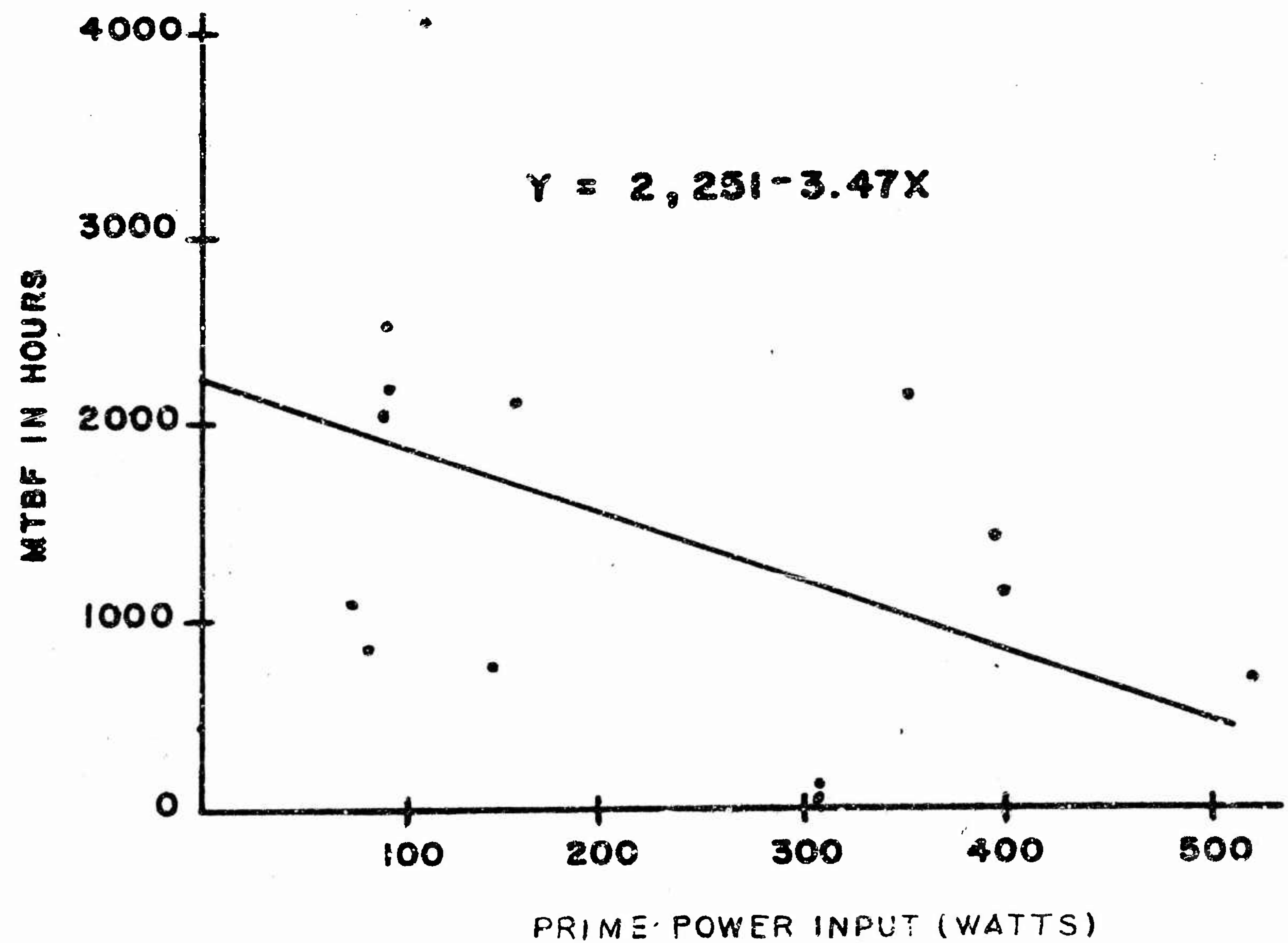


FIGURE 5. MTBF VS. PRIME POWER INPUT.

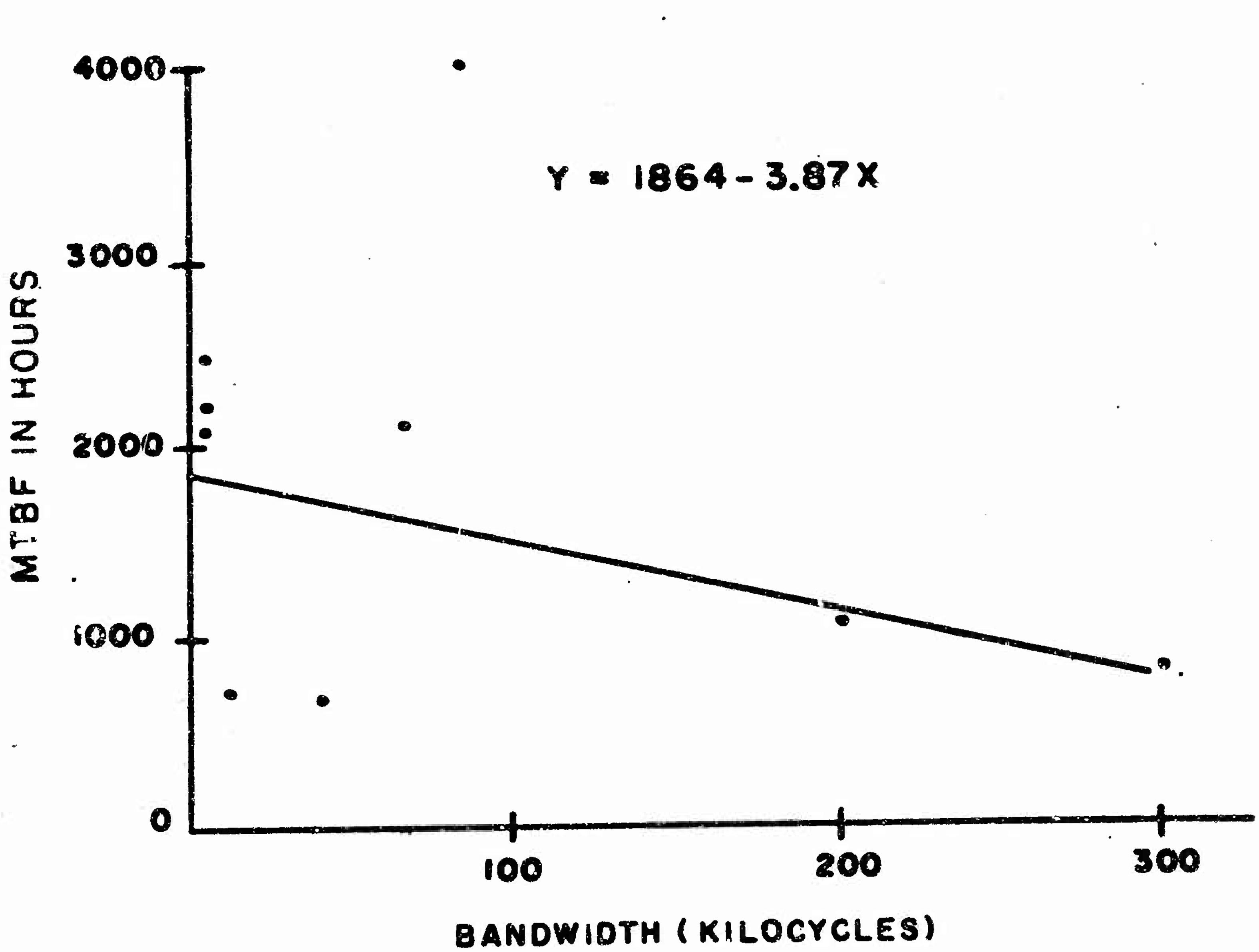


FIGURE 6. MTBF VS. BANDWIDTH.

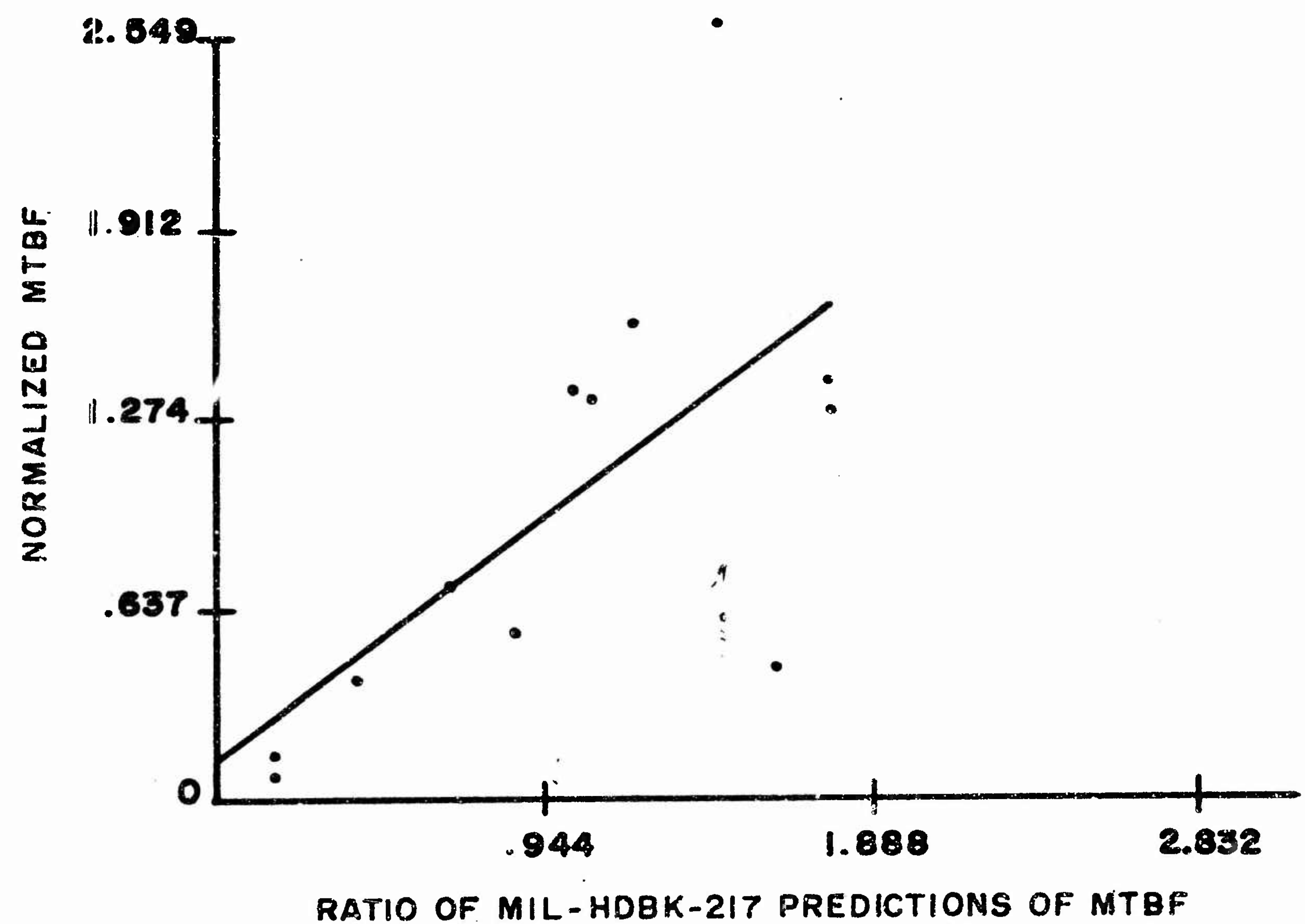


FIGURE 7. MTBF VS. RATIO OF MIL-HDBK-217 PREDICTIONS OF MTBF.

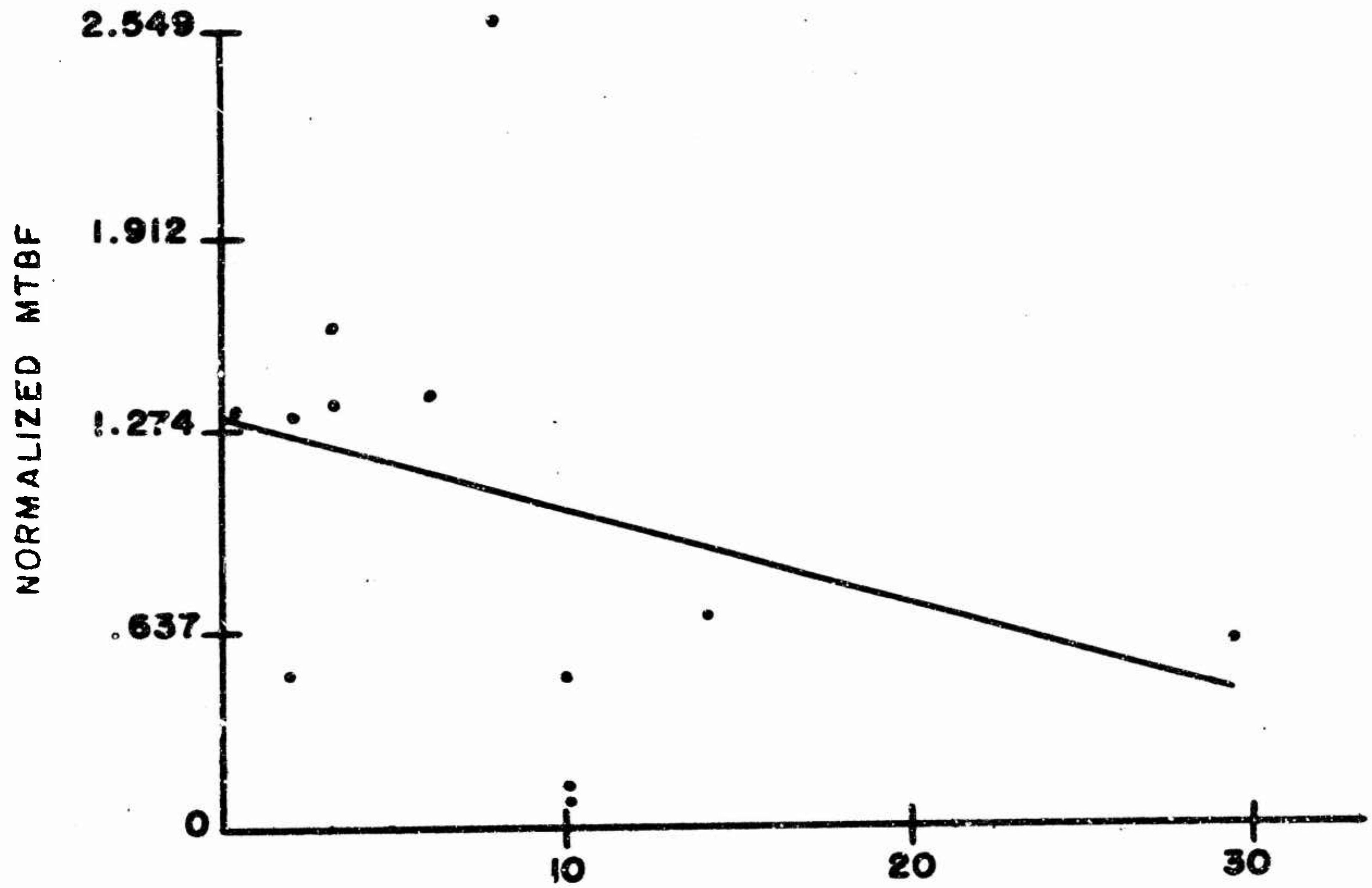


FIGURE 8. MTBF VS. RATIO OF SENSITIVITIES.

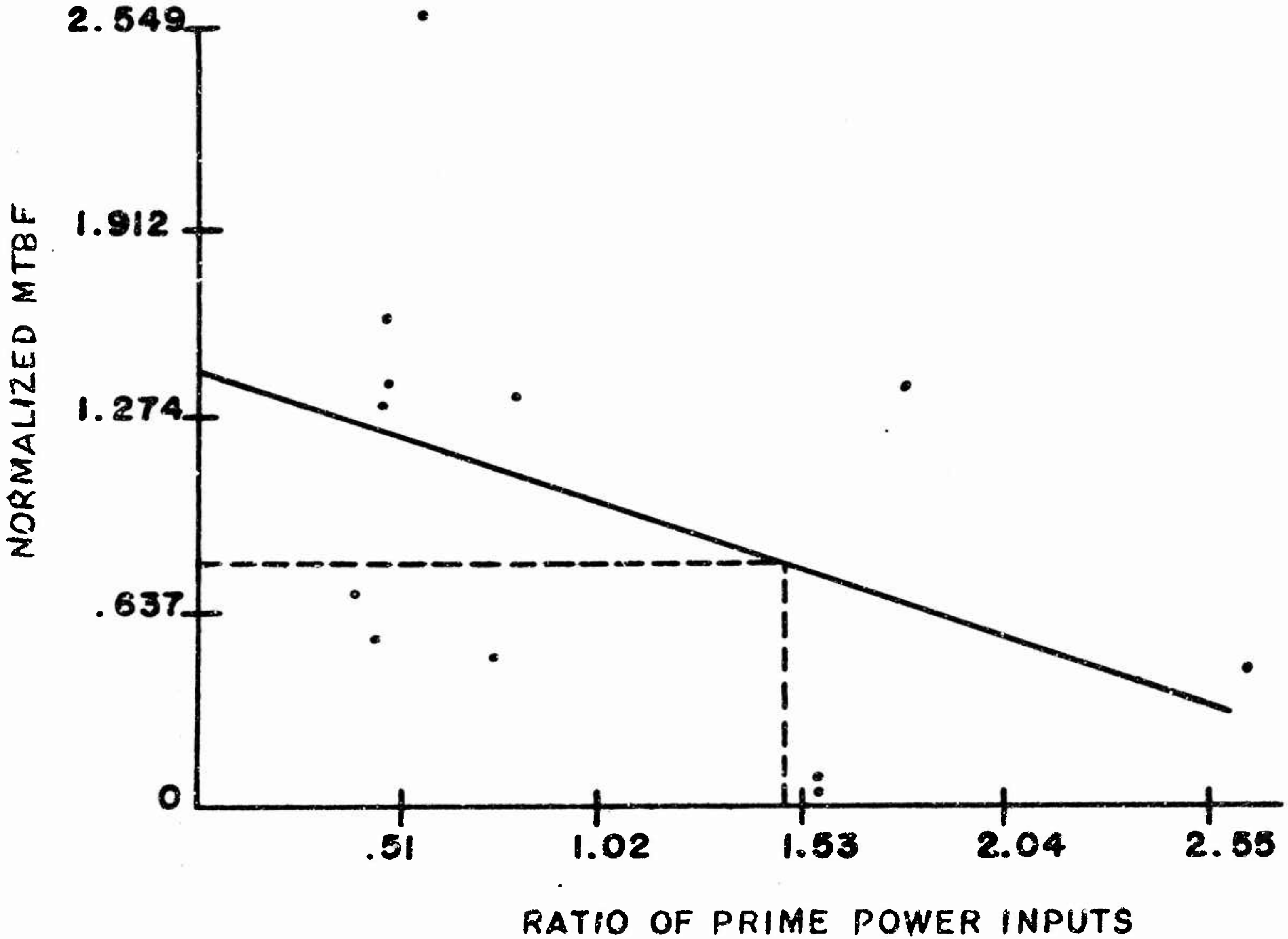


FIGURE 9. MTBF VS. RATIO OF PRIME POWER INPUTS.

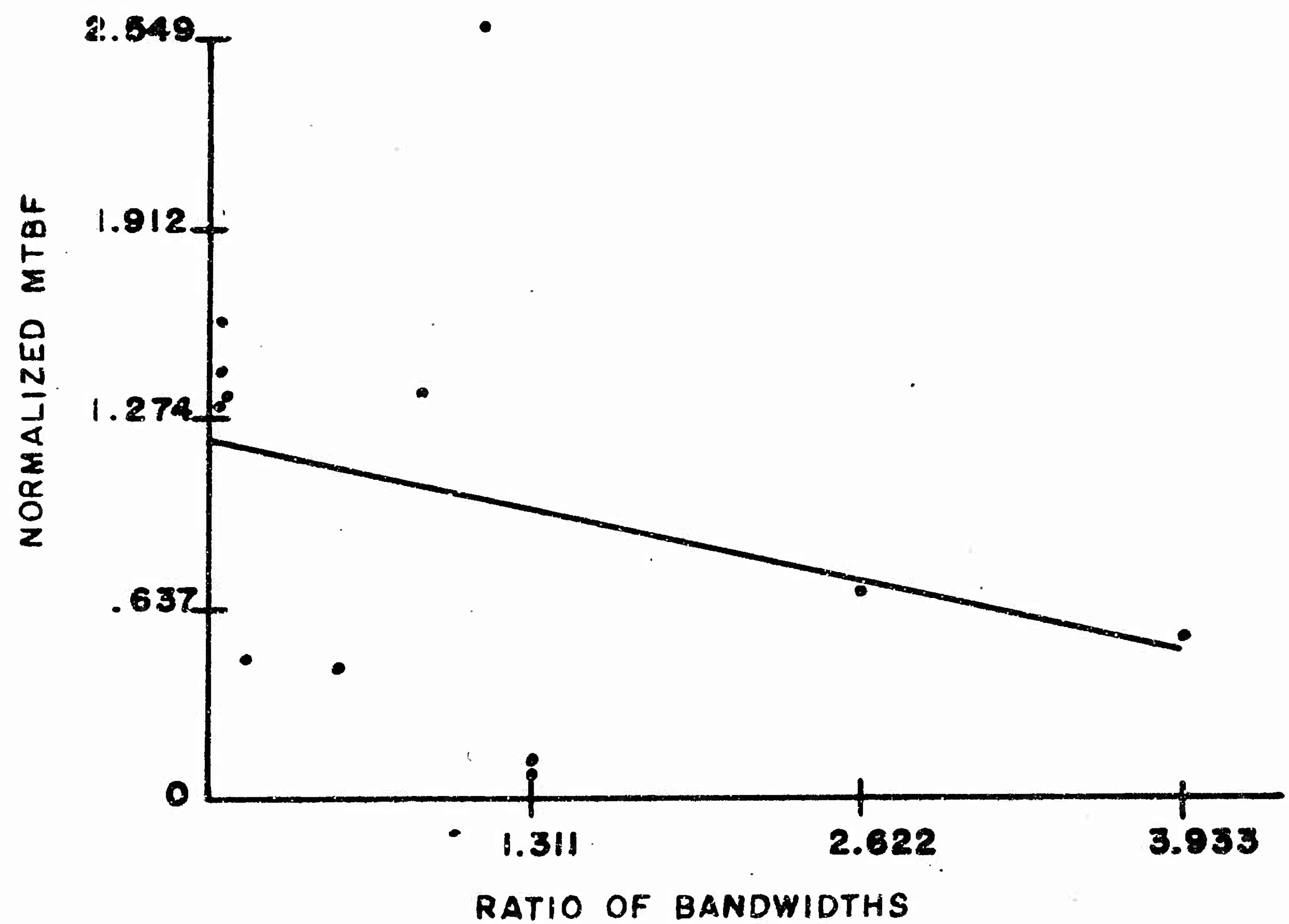


FIGURE 10. MTBF VS. RATIO OF BANDWIDTHS.

Purpose

To find a predicting equation for Y_i of the form $Y_i = b_0 + b_1X_{i1} + \ldots + b_pX_{ip}$, given n sets of observations on the p + 1 variables $(Y_i, X_{i1}, X_{i2}, \ldots, X_{ip})$ where where Y_i denotes the dependent variable, the X_i's denote the independent variables or predictors, and $i = 1, 2, \ldots, n$. The routine also calculates other related statistical quantities.

Methods

This routine uses the method of least squares; that is, the coefficients $b_0, b_1, \ldots, b_p$ are determined so that the quantity $\sum_i (Y_i - \hat{Y}_i)^2$ is minimized. $i = 1, 2, \ldots, n$.

Usage

In the following, $i = 1, 2, \ldots, n$
$j = 1, 2, \ldots, p$
$k = 1, 2, \ldots, p$

For complilation the subroutine must be preceded by two EQU cards setting the maximum number of observations and the maximum number of independent variables to be used in the program. These cards must be of the following form:

Loc.	Op.	Address
MAXP	EQU	(integer)
MAXN	EQU	(integer)

Calling sequence:

Loc.	Op.	Address, Tag, Decrement
L	TSX	(MLRG, 4)
L+1	PZE	Y1, c, d
L+2	PZE	p, 0, n
L+3	OCT	ABCDE
L+4	Error return	
L+5	Normal return	

Y1 Location of the first word of a block of storage reserved for the dependent and independent variables. The n (p+1) words of floating point data must be stored in successive locations as follows: $Y_1, Y_2, \ldots, Y_n, X_{11}, X_{21}, \ldots, X_{12}, X_{22}, \ldots, X_{n2}, \ldots, X_{1p}, X_{2p}, \ldots, X_{np}$.
The numbers in these locations are destroyed by the routine, but the Y_i's will have been restored before exit to the normal return if c=0 (see below).
c=0 or 1.

If c=0, the Y_i's will have been restored in locations Y1 through Y1 + n upon exit to the normal return.

If c=1, $(Y_i - \hat{Y}_i)$ will be stored in the above locations upon exit to the normal return.

d=0 or 1

If d=0, all variables, Y_1 through X_{np}, must be supplied.

If d=1, only variables Y_1 through X_{n1}, must be supplied, polynomial approximation is assumed, and appropriate powers of X_{n1} are calculated by the routine and stored in locations X_{12} through X_{np}. Adequate storage must be allowed for this.

p, The number of independent variables, or predictors.

n, The number of data points of the dependent variable (Y_1) set.

ABCDE Five octal digits representing a pattern of 15 binary bits which determine the information the routine will print. If a given bit is 1, the corresponding information will be printed; while if the bit is 0, that information will not be printed. The following is a tabulation of the 15 bits and the identification which will be printed with the corresponding data. Further definition and explanation of the information given are found in paragraph (A), Explanation of Output.

Bit			Identification
A:	4's bit	(1)	MEANS - YBAR, X1BAR, X2BAR, ETC. TO EPBAR
	2's bit	(2)	STANDARD DEVIATIONS
	1's bit	(3)	CORRELATION MATRIX PXP R(XJ, XK)
B:	4's bit	(4)	COLUMN CORRELATION VECTOR PX1 R(Y, XJ)
	2's bit	(5)	INVERSE MATRIX
	1's bit	(6)	STANDARD REGRESSION COEFFICIENTS
C:	4's bit	(7)	REGRESSION COEFFICIENTS
	2's bit	(8)	YHAT, Y, Y- YHAT, and X11, X12, ..., X1P.
	1's bit	(9)	CONFIDENCE INTERVAL
D:	4's bit	(10)	SUM OF RESIDUALS
	2's bit	(11)	MAX/Y - YHAT/
	1's bit	(12)	VARIANCE ST. D.
E:	4's bit	(13)	R SQUARED
	2's bit	(14)	F TEST
	1's bit	(15)	T TESTS (J = 1, 2, ..., P)

The programmer must define MAXP, MAXN, (OUT AND COMMON). An error return will occur for any of the following reasons:

(1) $n-p-1 < 0$.

(2) The confidence interval is imaginary.

(3) Any component of the correlation matrix is indeterminate, in which case INDETERMINATE will be printed in place of the correlation matrix.

(4) The correlation matrix is singular, in which case SINGULAR will be printed with the heading for the correlation matrix, and the correlation matrix and column correlation vector will be printed out, if requested, in the calling sequence. The correlation matrix will be called singular if $1 - r_{j\,jk} < 10 - 7$. The correlation matrix may be singular for other reasons, but these are not considered.

Storage Requirements

Subroutine: 1362 + 6 (MAXP) + (MAXN) + (MAXN) (MAXP) + (MAXP)^2
Common: 8 + 2(MAXP)

Exempt Symbols

(MLRG)

(A) Explanation of Output

In the following: $i=1, 2, \ldots, n$ $j=1, 2, \ldots, p$ $k=1, 2, \ldots, p$

The number in parentheses used below correspond to those printed as part of the identifying headings for the data which is printed out (see USAGE).

(1) The arithmetic means $\bar{Y}, \bar{X}_1, \bar{X}_2, \ldots, \bar{X}_p$. For example, $\bar{Y} = \frac{1}{n}\sum_i Y_i$

(2) The standard deviations of the quantities $Y, X_1, \ldots, X_p$. For example, the standard deviation of X_j is $\frac{1}{n-1}\sum_i (X_{ij} - \bar{X}_j)^2$.

(3) If $a_{jk} = \sum_i (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)$, $a_{j0} = \sum_i (Y_i - \bar{Y})(X_{ij} - \bar{X}_j)$.

$$a_{00} = \sum_i (Y_i - \bar{Y})^2$$

$$r_{jk} = r(X_j, X_k) = \frac{a_{jk}}{\sqrt{a_{jj}a_{kk}}} \qquad r(X_j, Y) = \frac{a_{j0}}{\sqrt{a_{00}a_{jj}}}$$

then the "normal equations of least squares" after an appropriate scaling, may be written as $[r(X,X)][b^*] = [r(X,Y)]$ where $[r(X,X)]$ is the pxp symmetric correlation matrix whose elements, r_{jk}, are printed.

(4) The elements of the px1 column correlation vector $[r(X,Y)]$, $[r(X_j,Y)]$.

(5) The elements of the inverse of $[r(X,X)]$, denoted by $[r(X,X)]^{-1}$.

(6) The standard regression coefficients, which are the elements of the px1 column vector $[b^*] = [r(X,X)]^{-1}[r(X,Y)]$ denoted by $b_1^*, \ldots, b_p^*$.

(7) The quantities $b_0, b_1, \ldots, b_p$ are the regression coefficients. The b_j are related to the b_j^* by the equation $b_j = b_j^*$

$$\sqrt{\frac{a_{00}}{a_{jj}}} \quad \text{and } b_0 = \bar{Y} - \sum_j b_j \bar{X}_j.$$

(8) $\hat{Y}_i, Y_i, Y_i - \hat{Y}_i$, and $X_{i1}, X_{i2}, \ldots, X_{ip}$. The $\hat{Y}_i$'s are computed from $\hat{Y}_i = b_0 + \sum_j b_j X_{ij}$.

(9) If one assumes that Y is normally distributed with variance σ^2, independent of the X's, then a $(1-\gamma)$ 100-percent confidence interval for the average value of Y corresponding to a given set of X_{ij}'s is given by $\hat{Y}_i + t\gamma$,

n-p-1 $\hat{\sigma}$ I_i where t_{γ} . n-p-1 (found in statistical tables). is the upper 100γ percent point of "Students" -t distribution with n-p-1 degrees of freedom, i.e., =Pr ($|t_{n-p-1}| < t_{\gamma}$, n-p-1). For example, if γ =0.05, n=25, and p=5, then $t_{0.05.19}$ = 2.093. The quantities printed are computed from

$$I_i = \sqrt{\frac{1}{n} + \sum_{j,k} c_{jk} (X_{ij} - X_j)(X_{ik} - X_k)}$$

where

$$c_{jk} = \frac{d_{jk}}{\sqrt{a_{jj}a_{kk}}}$$

and the d_{jk} are the elements of $[r(X,X)]^{-1}$.

(10) The sum of the residuals, $\sum_i (Y_i - \hat{Y}_i)$. Theoretically, this should be zero.

(11) The maximum residual, in absolute value.

(12) The variance, an unbiased estimate of σ^2 , is computed from $\hat{\sigma}^2 = \frac{1}{n-p-1} \sum_i (Y_i - \hat{Y}_i)^2$. If n-p-1 = 0, the quantity printed is $\sum_i (Y_i - \hat{Y}_i)^2$, which theoretically should be zero. The square root of $\hat{\sigma}^2$, the standard deviation, is also printed.

(13) The quantity R^2 is computed from $R^2 = 1 - \frac{\sum_i (Y_i - \hat{Y}_i)^2}{\sum_i (Y_i - \bar{Y})^2}$ where R is the coefficient of multiple correlation. The quantity R^2 measures the percentage improvement, using the predicting equation, relative to using the mean, $\bar{Y}$.

(14) The F-Test for testing the overall significance of the regression function is effected by computing $F = \left(\frac{n-p-1}{p}\right) \left(\frac{R^2}{1-R^2}\right)$ (which is printed) and comparing to F_{γ}. (p_1n-$p_2$1)(found in statistical tables) where γ = Pr($F_{n1,n2} F_{\gamma}$, (n_1, n_2 , n_1 =p, n_2=n-p-1. For example, if p=5, n=25, =0.05, F would have to exceed 2.74 to be considered significant at the 5-percent level. If the F-test is significant, one can then determine from the t-test which of the p predictors made it significant.

(15) The t-test for testing the hypothesis Bj=0, where Bj is the mean of b_j, is carried out by computing $t_j = \frac{b_j}{\hat{\sigma}\sqrt{c_j}}$ The t_j's are printed. If $|t_j|$, t_{γ}, n-p-1, then Bj is said to be significantly different from zero at the 100 percent level. If n-p-1=0, the F value printed will be zero and no t values will be given.

(B) CHECKS

1. The inverse matrix should be symmetrical.
2. The elements in the main diagonal of the inverse matrix should all be positive.

3. The sum of the residuals should be nearly zero.
4. The quantities I_i should all be greater than $\frac{1}{\sqrt{n}}$.

(C) REFERENCES

1. Ostle, B., "Statistics in Research," Chapter 8, The Iowa State College Press, Ames, Iowa, 1959.

2. Anderson, R.L. and Bancroft, T.A., "Statistical Theory in Research", Chapter 14, McGraw Hill, 1952.

APPENDIX II

USE OF REGRESSION EQUATIONS FOR DETERMINING WEIGHTING FACTORS*

The purpose of this appendix is to show that relative weights of various indices can be determined from the regression equations.

The form of the estimating equation is: $R_\gamma = R_B \sum_a I_a W_{a,\gamma}$ and the form of a regression equation is: $R = A_0 + A_1X_1 + A_2X_2 + \ldots A_pX_p$ where A_0 is a constant, the A_i's are constant coefficients, and the X's are the values of the indicators. To demonstrate the method of extracting the weighting factors from the regression equation, it is necessary to present tables of the coefficients of the regression equations. If there are p a's p such tables will be presented.

These p tables would represent the regression equation for the first available indicator, the first and second available indicator, the first, second, and third available indicator, etc. The coefficients in any one table will fulfill the function of providing the relative weights of each a used in the equation, while the particular equation used fulfills the function of weighting the availability of information on the indicators.

To derive the weighting factor it will be shown that the regression equations can be converted into the form of the estimating equation. Consider the case where there are a a's. There are then p! -1 regression equations.

$$R_1 = A_{0,1} + A_{1,1}X_1$$

$$R_2 = A_{0,2} + A_{2,2}X_2$$

$$\vdots \qquad \vdots$$

$$R_p = A_{0p} + A_{pp}X_p$$

$$R_{p+1} = A_{0,p+1} + A_{1,p+1}X_1 + A_{2,p+1}X_2$$

$$R_\gamma = A_{0,\gamma} + A_{a,\gamma}X_a + \ldots \text{ (all a in equation } \gamma)$$

$$R_{p!-2} = A_{0(p!-2)} + A_{1(p!-2)}X_1 + \ldots A_{p-1,(p!-2)}X_{p-1}$$

$$R_{p!-1} = A_{1(p!-1)} + A_{1(p!-1)}X_1 + \ldots + A_{p,(p!-1)}X_p$$

The general symbol for the number of a regression equation is γ and the general symbol for identifying a reliability indicator is a.

Let the first p equations be used to define R_B and I_a. Take $R_B = 1/p \sum_{i=1}^{p} A_{0,i}$ and define I_a as $I_a = \frac{A_{0,a} + A_{a,a}X_a}{R_B}$ $\quad a = 1, 2, \ldots p$

*Adapted from bibliography reference 12.

For the first p equations, let $W_{a,\gamma}$ be given by $W_{a,\gamma} = 1 \quad a = \gamma$
$W_{a,\gamma} = 0 \quad a \neq \gamma$

Then the first p equations can be written in the proper form, i.e.,

$$R_\gamma = R_B \sum_a I_a W_{a,\gamma} \qquad \gamma = 1, 2, \ldots p$$

Let us rewrite the regression equations for $\gamma > p$ in the following way:

$$R_\gamma = (A_{i,\gamma}X_i + A_{0,\gamma}/n) + (A_{j,\gamma}X_j + A_{0,\gamma}/n) + \ldots (A_{k,\gamma}X_k + A_{0,\gamma}/n)$$

where i, j, k... range over all the a's in the γ^{th} regression equation and n is the total number of a's in the γ^{th} equation, e.g.,

$$R_{p+1} = (A_{1,p+1}X_1 + A_{0,p+1}/2) + (A_{2,p+1}X_2 + A_{0,p+1}/2)$$

The regression equations now have one term for each reliability indicator that appears in the equation. If we let the weight be zero for the a's that are not present in an equation then it is possible to compare each term in the rewritten regression equation with a non-zero term in the linear weighted estimating equation, e.g.,

$$R_\gamma = (A_{i,\gamma}X_i + A_{0,\gamma}/n) + (A_{j,\gamma}X_j + A_{0,\gamma}/n) + \ldots \text{ and}$$

$$R_\gamma = R_B I_i W_{i,\gamma} + R_B I_j W_{j,\gamma} + \ldots$$

where again a takes on the values i, j, etc. If we replace $R_B I_a$ by $A_{a,a}X_a + A_{0,a}$ we have the following general formula for the weights:

$$W_{a,\gamma} = \frac{A_{a,\gamma}X_a + A_{0,\gamma}/n}{A_{a,a}X_a + A_{0,a}} \qquad = 0$$

a is present in γ^{th} regression equation
a is not present in γ^{th} regression equation

This completes the proof that the coefficients in the regression equation can be converted into the coefficients required for the estimating equation of the proposed technique.

BIBLIOGRAPHY

1. Bozovsky, Igor. Reliability Theory and Practice, Prentice Hall Inc, Englewood Cliffs, New Jersey. 1961.

2. Calabro, S.R. Reliability Principles and Practices, McGraw Hill Book Company, Inc., New York. 1962.

3. Chorafas, Dimitris N., Dr. Statistical Processes and Reliability Engineering, D. Van Nostrand Company, Inc., Princeton, New Jersey. 1960.

4. Current Reliability and Maintainability Reports, Rome Air Development Center, Griffis Air Force Base, Rome, New York. December 1963.

5. Dummer, G.W.A. and Griffin, N. Electronic Equipment Reliability, John Wiley and Sons Inc., New York. 1960

6. Freund, John E. Mathematical Statistics, Prentice Hall Inc., Englewood Cliffs, New Jersey. January 1963.

7. Hoel, Paul G. Introduction to Mathematical Statistics, Third Edition, John Wiley and Sons Inc., New York. 1962.

8. Lloyd, David K. and Lipow, Myron. Reliability: Management, Methods, and Mathematics, Prentice Hall Inc., Englewood Cliffs, New Jersey. 1962.

9. Methods for Predicting Combined Electronic and Mechanical System Reliability, IIT Research Institute, Chicago, Illinois. 30 September 1963, June 1965.

10. Pieruschka, Erich. Principles of Reliability, Prentice Hall, Inc., Englewood Cliffs, New Jersey. 1963.

11. Proceedings Fourth Through Tenth National Symposium on Reliability and Quality Control, Sponsored by ASQC, Washington, D.C.. January 1958 - January 1964.

12. Program to Establish Review - Point Criteria for Reliability Monitoring, Hughes Aircraft Company, Fullerton, California. July 1964, January 1965.

13. Reliability Abstracts and Technical Reviews, Serial Numbers 1 - 1635, National Aeronautics and Space Administration, Washington, D.C. 1961 - 1964.

14. Reliability Theory and Practice, Arinc Research Corporation, Washington, D.C. 1962.

15. Roberts, Norman H. Mathematical Methods In Reliability Engineering, McGraw Hill Book Company, New York. 1964.

16. SAE-ASME-AIAA Conference Proceedings, Aerospace Reliability and Maintainability Conference, Washington, D.C. July 1964.

17. Sandler, Gerald H. System Reliability Engineering, Prentice Hall Inc., Englewood Cliffs, New Jersey, 1963.

18. System Reliability Prediction by Function, Arinc Research Corporation, Washington, D.C. August 1963, March 1965.

19. System Reliability Prediction by Function, Federal Electric Corporation, Paramus, New Jersey. May 1963, March 1965.

20. Tommerdahl, J.B. and Nelson A.C. Design Stage Reliability Prediction for Electronic Equipment, Durham, North Carolina. October 1963.

21. United States Department of Defense, MIL -HDBK - 217 Government Printing Office, Washington, D.C. 8 August 1962.

A MODEL FOR EVALUATING AND SELECTING FUTURE
ARMY FUEL SYSTEMS

by
Mr. Sidney Sobelman
Office of the Chief of Engineers
Engineer Strategic Studies Group

ABSTRACT: The concept of a Nuclear Power Energy Depot (NPED) is being studied to meet the Army objectives of reducing the long logistics tail and increasing the mobility of the field Army. The goal is improved cost-effectiveness, where effectiveness implies combat effectiveness. Thus, a mathematical evaluation model should indicate how alternative fuel or energy systems are related to the efficiency of an Army's response rate and mobility in completing a mission. The possible effect of an NPED might be as revolutionary as a change from mule to truck transportation.

This paper explores only the problem of establishing a decision model into which data produced by engineering studies and scenario plays will be inserted. An earlier classified paper (Ref. 1) on the subject of the future Army's fuel or energy system was presented at the 1964 Army Operations Research Symposium. This has been a continuing study by the Engineer Strategic Studies Group. The concepts and objectives are now reexamined for interpretation into quantifiable cost-benefit or cost-effectiveness measures for total system analysis.

The concept of a mobile, compact nuclear reactor (MCR) which can move forward with the combat forces and be relatively independent of long logistic lines is in competition with cheap and efficient petroleum (POL) systems, which are already established. The nuclear reactor has certain alternatives in its own design, size, and characteristics and also in its type of energy output, specifically, either chemical fuels like ammonia and hydrogen (extracted from air and water) or direct electrical output. The Army aircraft vehicles using the output of the Nuclear Power Energy Depot may have turbine engines, fuel cells, or electrochemical energy storage systems (EESS). Associated with these devices, a vehicle system may have a mechanical or hydraulic drive or either a single electric motor drive or individual wheel electric motors. For evaluation of all feasible alternatives, the technology of the 1975-1980 time frame is to be considered. An unclassified

technical description of the Nuclear Energy Depots and using devices is given in Reference 16.

Table I indicates the three mobility classes of the NPED that are being studied. The mobility Class I package of 35,000 pounds is preferred for greatest ease of transportation over fair roads and across 20-ton bridges, but the mobility Class III NPED has almost four times the output of the Class I depot. The Class III NPED's travel however, will be more restricted but not more so than that of a tank remover plus tank, which weigh 100,000 pounds. Almost four times as many Class I depots will be needed as Class III depots to produce the same amount of energy, but the Class I depots can be moved more easily and more closely to the FEBA (Front Edge of the Battle Area). A Class I or III depot may cost about the same in research and development, production, and operation and maintenance.

TABLE I
NPED MOBILITY CLASSES

Class	Max. Wt/Unit (lb)	Max. Output MW (electric)	Equiv Gal POL 1/
I	35,000	6.3	5,360
II	65,000	14.7	12,500
III	100,000	23.7	20,000

1/ Using an equivalence of .75 KW/gal POL and a double transfer of energy efficiency of (0.8)(0.8) = 0.64. Thus, multiply megawatts by (1000)(.64)/.75 or 850 to obtain equivalent gallons.

Table IIa lists the alternatives being considered. The one alternative previously considered but no longer a candidate is the I.C. (internal combustion) engine using ammonia as a fuel. Its elimination is not based on ability to use ammonia in an I.C. engine (as can be done and even with engineering efficiency equivalent to that of POL if a supercharger is used), but primarily by the time frame which, with very little doubt, can permit the development of the more efficient fuel cell using ammonia. Other considerations negate the ammonia-I.C. engine such as the larger fuel tank for ammonia to get equivalent mileage or running time; about 2½ times greater weight of ammonia is needed. Table IIb is a rearrangement of Table IIa to consolidate the fuel or energy input possible for each user equipment.

TABLE IIa

Altn Fuel or Energy Input A	Intermediate Field Storage B	Associated Field Conversion C	1975-80 Users 2/ —
1. POL a. Mogas b. Diesel c. Avgas d. Jet fuel	1. 5-gal can (Jerrican) Collapsible tank Truck tank	1a-none needed 1b-Field I.C. electric generator 1/	A1/B1/C1a-Turbine engine aircraft A1/B1/C1b-EESS 3/ vehicle 1/ A1a/B1/C1a-I.C. engine vehicle
2. Ammonia, liq, anhyd	2. Truck tank	2a-none needed 2b-Fuel cell van mounted	A2/B2/C2a-Fuel cell vehicle A2/B2/C2b-EESS vehicle A2/B2/C2a-Turbine engine aircraft
3. Hydrogen, liq or sludge	3. Truck tank, special, insulated	3a-none needed 3b-Fuel cell van mounted 3c-Turbine generator	A3/B3/C3a-Turbine engine aircraft A3/B3/C3a-Fuel cell vehicle A3/B3/C3b-EESS vehicle A3/B3/C3c-EESS vehicle
4. Electricity	4a. EESS storage van 4b. "Jerrican" replacement module	4. none needed	A4/B4a/C4-EESS vehicle A4/B4b/C4-EESS vehicle

1/ This would be an emergency or slack equipment where POL is indigent, e.g., CONUS.

2/ All ground vehicles are expected to have electric drives, except perhaps I.C. engine vehicles.

3/ EESS denotes an Electrochemical Energy Storage System, consisting of a regenerable electrochemical cell (a secondary battery).

TABLE IIb

1975-80 User	Fuel or Energy Input	Supply Sources
Current I. C. engine vehicle	POL	POL supply line
Fuel Cell Vehicle (electric drive)	1. Ammonia 2. Hydrogen	1. NPED/Ammonia Plant 2. NPED/Hydrogen Plant
EESS Vehicle (electric drive)	Electricity	1. NPED direct output 2. EESS intermediate storage van 3. "Jerrican" replacement module 4. Fuel cell intermediate storage van (NH_3 or H_2) 5. Field turbine generator (multifuel) 6. Field I.C. generator (emergency)
Turbine Engine Aircraft1/	1. POL 2. Ammonia 3. Hydrogen	See above.

1/ A regenerative engine once designed to burn hydrogen might easily be used for POL or ammonia, provided the fuel tanks are properly purged or exchanged.

A brief word about the 1975-80 time frame would not be remiss. It is expected that the clear cost-effectiveness advantage of certain technological developments will introduce them by 1980 with or without the development of the NPED. It is therefore reasonable to exclude their costs from the cost-effectiveness study of an NPED. On the other hand, if the NPED is to be developed and procured, coordination of and influencing the direction of development of these other items, listed below, can increase overall Army efficiency (i.e., avoid harmful sub-optimizations):

1. Electric drives on vehicles. Advances in electric motor design and controls indicate both in government and industry that the number of automotive parts can be decreased, efficiency increased, and maintenance

and operating costs decreased. The efficient, electric-producing fuel cell and EESS then becomes natural candidates to replace an I.C. engine and generator. (See Reference 15 which is basically valid but needs technological up-dating.) Ammonia and hydrogen are very efficient fuels for the fuel cell and electricity for the EESS is obtainable from electric outlets (but must be rectified from AC to DC) or electric generators or convertors.

2. Fuel cells in commercial/household use, space and vehicles. As a prime example, a study has been initiated for developing a fuel cell to provide the electrical needs of a household; natural gas (after scrubbing out catalytic "poisons"?) which contains hydrogen, might be the fuel. Reliable use of fuel cell burning hydrogen has been demonstrated in space flights. For automotive use, see Reference 15.

3. EESS drives for special vehicles. Golf carts, industrial delivery vans and lift forks, hospital motorized chairs, and the like, now exist with the conventional types of storage battery. See also Reference 15.

4. Vastly increased industrial ammonia production. The doubling of ammonia output capacity by 1970 is expected. The more efficient processes might lead to decrease in prices. This portends well since a field army type of vehicle, as one using amonia, must still be useable under peacetime or CONUS conditions.

5. Military rapid deployment capability. The C-5A transport plane and the RDL (Rapid Deployment Logistics) vessel will be available in the 1970's for a maximum 30 day build-up or response time to crisis in any part of the world. Althoug POL may be plentiful for initial deliveries to overseas ports, the tonnage of planes, cargo vessels, storage tanks, pipeline material, tank trucks, barges, and railroad tanks for inland movement of bulk or packaged POL is a drag and hindrance to the total rapid logistics delivery and supply line problem.

A totality of analysis must consider: Whole systems to avoid harmful sub-optimization; the meeting of the Army's objectives rather than those of organizational subdivisions; technical feasibility and engineering efficiencies without excluding the costs of RDT&E, investment, and operations and maintenance; and, the value of time in quicker responsiveness or build-up to combat effectiveness and in faster conclusion of a total conflict. For all of this, the best estimates of knowledgeable individuals and groups must be and are being obtained and validated even though the hardware systems may not yet have been designed, let alone manufactured.

A brief background of the problem must be presented in order to comprehend fully the mathematical modeling which will follow. First, the claim has been made that a World War III conflict would probably require twice the amount of POL used in World War II. Also, consider that in World War II and Korea about 60 percent by weight of all supplies shipped consisted of POL. The basic usage demand for POL is compounded by the vulnerability of long lines of communications from oil fields to refineries and storage points to overseas bases and ports and thence to Army Service Areas in the Combat Zone by rail, barge, pipeline, truck, and aerial delivery and drop. To assure an acceptable level of 2-3 1/ days of POL supply in the combat zone, supply depots or divisional units, in general 42 days of supply back this level up for a total of 45 days of supply in the Theater of Operations (see Figure 1). Thus, it is natural to consider at least a partial substitution for POL with another type of fuel or energy that can be produced wholly in the Theater of Operations, particularly in the combat zone, as can be feasibly done by an NPED.

We take no issue here about the claims of limited supply or depletion of world quantities of POL. Obviously there must be some limit to the underground sources of petroleum, but then there are great potential sources of POL, as in recovery from oil shale, which promise the extended supply of POL for many, many years. Complete substitution for POL is not necessarily expected nor is it necessary in order that an alternate or substitute fuel or energy source be useful or desirable. As part of the problem, unless other fuels, like hydrogen to be burned in gas turbines, are developed for Air Force and Army aircraft, POL will have to be continued to be used; but most or all ground vehicles, not having the air-lift -weight problem, can feasibly operate on an alternate fuel system.

The previous Engineer Strategic Studies Group paper (Ref. 1) used as a measure of effectiveness of different fuel systems the reduction of the total tonnage moved in moderately large Theaters of Operation. Costs were not considered, but were reserved for the next or present phase where six smaller, relatively isolated scenarios have been selected for play. The details of scenario play generate fuel or energy demands and costs. In the previous study, however, the plotting of cumulative tonnage against time after D-day does provide a picture of potential value for alternate systems.

1/ Reference 12 recommends that field commanders obtain a minimum of 2-3 days' supply for an offensive operation and 3-5 days' supply for a defensive operation.

Sutherland (Ref. 2) discusses the advantages of the scenario-play approach, which is essentially a one-sided game. The advantages expected are:

1. To reveal or permit examination of certain kinds of inter-relations or interactions.

2. To uncover missed details in new concepts.

3. To give insight into new tactical possibilities.

4. To reveal important areas of ignorance.

5. To point up any difficulties.

We also expect the following advantages from exercising the scenarios selected:

6. A representative variety of environments and conditions are included.

7. Each scenario serves as a "standard". The effect of the "expertness" of the team, good or bad, remains constant.

8. Fuel or energy demand profiles, with its peaks and valleys, will become known in detail.

9. Magnitude of output and number of the three different size fuel producers (mobility classes) can be tested out in different locations and different lengths of time between moves.

10. Numbers of people and equipment needed to support a field effort can be estimated accurately.

11. Fewer replications and lower study costs are involved than in using a "stochastic" two-person game, yet the detail of hourly and daily changes in mode and environment can be treated.

12. Costs involved in the use of alternative systems can be collected fairly actively in several ways including by the current budgetary and program package requirements.

The disadvantage of the scenario play, which is a "deterministic" process, lies in its inability to show directly the effect of different energy systems on the relative combat effectiveness of an organic military

unit. It is hoped, though, that the data generated by scenario play and subsequent data summarizations, can be used in evaluation models mirroring the Army's objectives.

In our approach, the scenario-play holds constant the established effectiveness, the ability to meet a fuel or energy demand; cost-generating activities of the various alternative systems create different system costs. The other general approach is to hold the system cost constant and determine the varying effectiveness of the alternatives (Refs. 7, 8, 9), not a very practical approach in this problem. In discussing cost-gain (or cost-benefit or cost-effectiveness) analysis, McKean (Ref. 8) cautions against the use of ratios like that of gains to costs. He says that good (though never perfect) criteria take the form of: (1) maximum gains minus costs (wherever possible), (2) maximum achievement of an objective for a given cost, or (3) minimum cost of achieving a specified amount of an objective. A variation of criteria (1) will be followed herein.

The most direct and simplest approach here to an evaluation model might be to consider the set of operations as an inventory or supply model in relative isolation from tactics and strategy, the methods "to gain, expand, or protect a nation's resources--whether those resources are persons, ideas, or materials (Ref. 10)". As an inventory model, the essential parameters would be: quantities, replacement and lead times, average rates of input and demand, and the costs of maintaining inventory, set-up or replacement, and shortages; the measure of effectiveness would then be to find the minimum cost system involved in meeting a fuel or energy demand.

A review of a major war effort (Ref. 11 for World War II) emphasizes, however, that supply, particulary of fuel (POL) and vehicles operating on it, is not inseparable from tactics and tactical planning, since "supply, instead of holding her rightful position as the handmaiden of battle, could become war's mistress". After the Normandy invasion and by the time of September 1963, COMMZ was unable to meet the demands placed upon it, and General Eisenhower halted offensive operations on a large part of the front and concentrated the bulk of the Allied resources on a relatively narrow front in the north.

The analyst soon recognizes, as in Reference 11, that the logistics problem involves <u>a rate of build-up</u> and <u>capability of support</u>, and that the latter can strongly be effected by <u>unexpected</u> accelerations, as from D-day to D+79 in WWII, and by <u>unexpected</u> delays. Also, when lines of communications (LC's) are short, we can adjust to the unexpected or the uncertainties of nature more easily than when LC's are long. CDOG (Ref. 13) then wishes to: increase mobility, increase response or reaction time,

to cope with a conflict or emergency situation in any part of the world, reduce the long logistics tail, and reduce the tonnage of Army supplies and equipment. Add to these the need to increase firepower capability and improvement of tactics (methods of operation, including degree of dispersion), but which are not to be tackled in detail herein.

In these terms, the NPED concept does not of itself increase the capability of any vehicle to move any faster, but system mobility, here defined as rate capability of an army to expand its area of control, may be increased. The NPED should be capable of delivery to a field army supply point (a key point at which to compare costs and delivery rates of the alternative systems) by C-5A transports and set up for operation in a matter of a few days. The long logistics tail need no longer be continuous but might consist of NPEDs in BASEC and ADSEC of COMMZ, in Field Army Service Area (FASA) depots, and at FASA or Corps Area Supply Points; interarea or intersectional deliveries of fuel would not be needed. The tonnage of stored reserves of fuel, equipment for storage tanks and pipelines, and of interarea POL delivery trucks, barges, railroad cars could be drastically reduced. The supply of fuel or energy becomes very much more controllable or deterministic, rather than stochastic, to deal with the "unexpected" accelerations. Unexpected delays directly related to the length of the POL logistics tail, as due to sabotage, "black market", submarine warfare, availability of transportation, vulnerability to enemy attack and ambush, adverse weather, etc., may well be reduced. The cost of prepositioning POL and its storage and pipeline facilities in many parts of the world might be obviated. We also need to know whether the Corps of Engineers, Quartermaster and Transportation Corps personnel involved in a POL line of communications will be more or less than the specialized personnel needed to operate and maintain an NPED and deliver its fuel or energy.

Whether the Field Army Supply Point is at a point of infiltration ready for expansion (e.g., at a beachhead) or is at a point reached by D+30, D+60, etc., it is a main supply distribution point to the divisional forces which do the fighting and advancing. It is normally far enough to the rear of FEBA not to come under direct attack by enemy field artillery. This is the key point where the nuclear energy depot might be delivered by a C-5A transport plane to support the Division Area. See Figure 1.

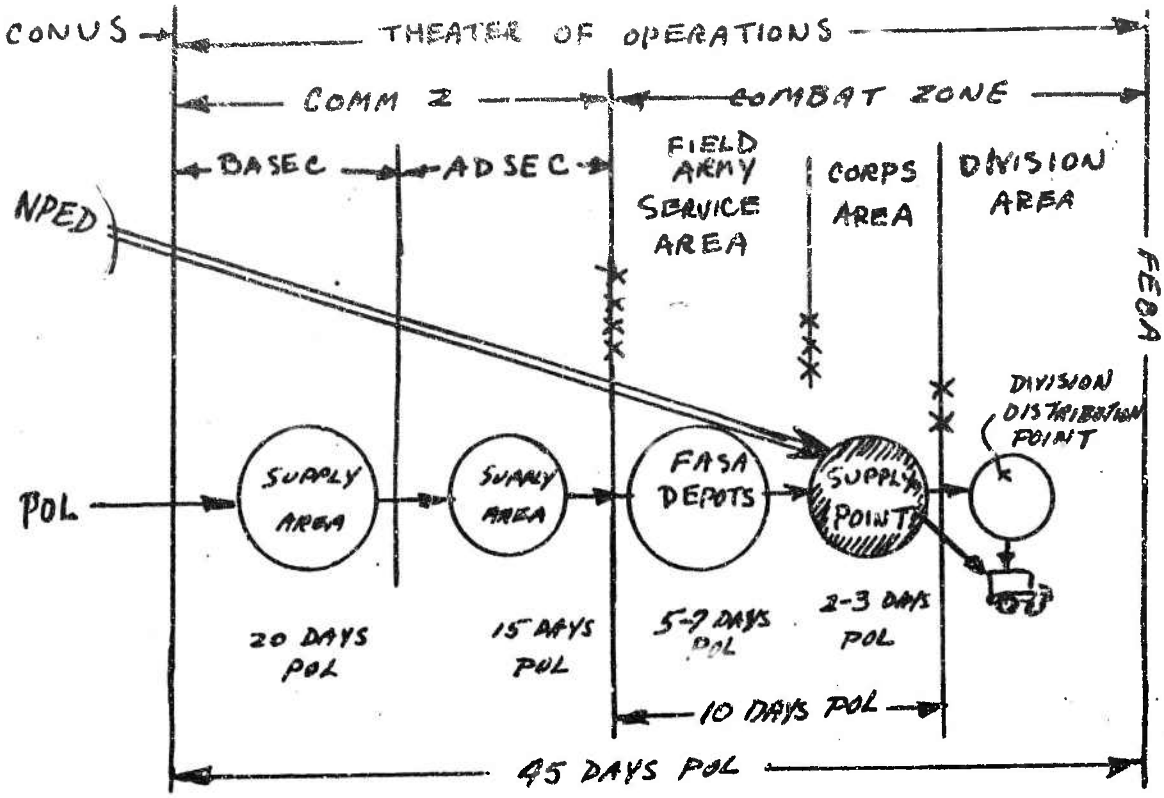

CONUS
THEATER OF OPERATIONS
COMM Z
COMBAT ZONE
BASEC
ADSEC
FIELD ARMY SERVICE AREA
CORPS AREA
DIVISION AREA
NPED
FEBA
DIVISION DISTRIBUTION POINT
POL
SUPPLY AREA
SUPPLY AREA
FASA DEPOTS
SUPPLY POINT
20 DAYS POL
15 DAYS POL
5-7 DAYS POL
2-3 DAYS POL
10 DAYS POL
45 DAYS POL

FIGURE 1.

Many "benefits" or "pay-offs" have just been mentioned but we recognize that each is attainable only at a cost. To set up an evaluation model in terms of costs and benefits (effectiveness), we will need to define terms and concepts. We need also to relate mobility, firepower, and dispersion, per References 5 and 4, even though firepower for this study is "fixed" by weapon and ammunition characteristics and the tonnage delivered is for use by a specified or planned number of combatants. Finally, there must be some greater value for higher rates of response or of mobility or in terms of time, shorter leadtime or response time or a shorter total war, as discussed and modeled in References 3 and 4. Imbedded in the total combat model is a fuel or energy demand created by the need to move tonnage and men as well as to move the fuel or energy source itself.

Consider that a threat exists. The decision to cope with it militarily and successfully is made. Alternative plans are postulated, each plan having associated tactics and logistics to overcome the restrictions imposed by the enemy, weather, terrain, human and equipment capabilities, etc. Relying on experience (e.g., Reference 12) game theory and gaming, and Lanchester's equations (modified or not), the number of men, N_1, and the tons and kinds of weapons and ammunition, S_1, needed in the Divisional Area, per man (or group of men) are first stated. Data are expected to be expressed for a Division Day and related to tons and to men.

Pay-off, effectiveness, or the "benefit" of all operations is the ability to expand its tonnage at a certain rate until the Total Area of Interest is in full control. Total Area of Interest divided by an average rate of expansion of our Area of Control will provide a calculated number of days of combat. The "gross" effectiveness, however is obtained at a cost, which when subtracted from a dollar equivalence of the gross effectiveness will provide a "net" effectiveness value.

These various parameters, predetermined by the military plan or variables in our study, are defined and symbolically represented as follows:

Z - Value or worth of the alternative system considered. Unit: Dollars

D - Dollars/ton for set up of a division.

R_1 - Rate of expansion of division combat forces' area of control. Initial unit: sq miles/day. Arbitrarily, one can say that

its dollar value is at least equal to its initial fixed set up costs for men, weapons, and equipment plus the variable costs for rations, fuel, and ammunition to maintain its rate of expansion capability for 10 days.

T - Days of total combat.

C_1 - Cost per day per ton per man of expenditures which are due to losses or consumption needing replacements, reserves, and spares to bring strength up to original. Included are salaries on the basis that the men are recruited only because the conflict exists.

C_2 - Cost per day per man-ton-mile of moving, storing, and maintaining tons of replacements, reserves, and spares in the support areas up to the Army Supply Points. If prepositioning is involved, the equivalent cost of moving these tonnages to the prepositioned points should be used.

C_3 - Cost per mile per ton-day, prior to D-day, to move tons of men, materiel, and supplies up to initial point of expansion (initial Supply Point).

t - Response time, which for RDL ships and C-5A transportation would be from D-30 to D-day.

C_4 - Wartime price incremental costs, per ton per day, if applicable, due to rates of demand exceeding efficient peacetime supply rates, e.g., POL going from .15/gallon to .22/gallon.

P_1 - Number of divisional combat area men.

P_2 - Number of support personnel.

S_2 - Tons of weapons, combat vehicles, and equipment/division/man.

S_5 - Tons of ammunition/division-day/man

S_1 - Tons of rations/division-day/man

S_3 - Tons of fuel/division-day/man

S_4 - Tons of non-combat vehicles in combat areas.

S_6 - Tons of non-combat vehicles in support area.

L - Miles

First, the value or worth of a system is:

$$(\text{Value}) = \begin{Bmatrix}\text{Gross}\\\text{Pay-Off}\end{Bmatrix} - \begin{Bmatrix}\text{Combat}\\\text{Costs}\end{Bmatrix} - \begin{Bmatrix}\text{Support}\\\text{Area}\\\text{Costs}\end{Bmatrix} - \begin{Bmatrix}\text{Response}\\\text{Time}\\\text{Costs}\end{Bmatrix} - \begin{Bmatrix}\text{Wartime}\\\text{Price}\\\text{Increment}\end{Bmatrix}$$

$$Z = D\left\{P_1(S_5)\cdot T+P_1S_2\right\} - C_1P_1(S_1+S_3+S_4+S_5)T-C_2P_2(S_1+S_2+S_3+S_5+S_6)L\cdot T$$
$$-C_3(S_1+S_2+S_3+S_4+S_5)L\cdot t - C_4S_3\cdot T \qquad (1)$$

The total combat time, T, is determined by the average rate of area expansion (R_1) and total area (A_T) to be covered, so that:

$$\frac{A_T}{R_1} = T \text{ days} \qquad (2)$$

The expression $\left\{P_1(S_5)\cdot T+P_1S_2\right\}$ contains the firepower capability so that an increase in the number of combat men and weapons and/or ammunition per man increases the firepower for the same number of days. Also, an increase in the rate, R_ϕ, would decrease the number of days of combat. If, perhaps, less trucks, S_5, were needed in support area to carry fuel and these were assigned to carry more men and ammunition, R_1 might be increased and the combat time, T, could be reduced in several parts of the equation; a higher value Z would result.

Another addition to the model will have to be made which represents the additional days of combat resulting from the compound probability that an unexpected delay or loss in the logistics line occurs simultaneously with the occurrence of a peak demand in the combat zone. It is expected that the probability of an unexpected loss is directly proportional to the length in miles of the lines of communication and the fraction of coverage, protection, or surveillance of the LOC to deter losses. If this fraction is f, a tentative probability model is:

$$P(\text{no loss}) = (1-f)e^{-f} \qquad (3)$$

$$P(\text{loss}) = 1-(1-f)e^{-f} = 1-e^{-f} + fe^{-f} \qquad (4)$$

$$P(\text{no loss}) + P(\text{loss}) = 1.00 \qquad (5)$$

REFERENCES

1. "An Operational Analysis of Nuclear-Powered Energy Depot Systems for the Army--A Potential Revolution in Fuel Logistics (U)," by George H. Orrell, US Army O.R. Symposium, May 1964. Proceedings-PartII.

2. "Some Basic Problems in War Gaming of Ground Combat," by Wm H. Sutherland, RAC-TP-153, March 1965.

3. "A Modern Dynamic Approach to Product Development," by S. Sobelman, 1959, OTS, Dept of Commerce, AD 209,505.

4. "Game Theory, Gaming, and Simulation," by S. Sobelman, US Army O.R. Symposium, March 1963, Proceedings-Part I.

5. "Historical Trends Related to Weapon Lethality, "Historical Evaluation and Research Organization (HERO), Volumes I, II, and III, 15 October 1964.

6. Blank.

7. "Military Cost Analysis," by Robert N. Grosse and Arnold Proschan, RAC-P-6, August 1965 (AD 622,113).

8. "Economics of Defense," by Roland N. McKean, Rand P-2926, July 1964.

9. "Cost-Effectiveness: Estimating System Costs," J. D. McCullough, Rand P-3229, September 1965.

10. "Combat Service Support," USACGSC RB 101-10-2, US Army Command and General Staff College, Ft. Leavenworth, Kansas, 15 June 1963, page 10.

11. "European Theater of Operations, Logistical Support of the Armies," Volumes I and II, by Roland G. Ruppenthal, Washington, 1953.

12. "Department of the Army Staff Officers' Field Manual. Organization, Technical and Logistics Data," FM 101-10, 3 Parts, October 1961.

13. "Combat Development Objectives Guide (CDOG)," USA Combat Development Command.

14. "Interaction of Firepower, Mobility, and Dispersion," by Major Wm. G. Stewart, Military Review, pp. 26-33, March 1960.

15. "Electric Motor Cars," by George A. Hoffman, Rand Memorandum RM-3298-FF, March 1963 (AD 615,883).

16. "Energy Depot Concept," SP-263, published November 1964 by SAE.

THE SIMULATION AND EVALUATION OF TRAFFIC LOADS AND COMMUNICATIONS NETWORKS IN A THEATER AREA

Mr. Theodore N. Truske
Combined Arms Research Office
Booz, Allen Applied Research Inc.
Fort Leavenworth, Kansas
Cognizant Agency: U.S. Army Combat Developments Command

I. PROBLEM

The design of a Theater Army Communications System (TACS-70) 1/ is to be based upon a survey of theater communications user requirements. In addition cost-effectiveness relationships are to be developed which can be used to design a minimal cost theater communications system. This paper describes the activities related to:

- Obtaining valid, theater-communications-user requirements data,
- Using the requirements data to determine traffic loads in various, proposed communications networks and,
- Describing the traffic behavior in terms useful in cost-effectiveness evaluations.

II. APPROACH

The major steps in satisfying the problem requirements include the following:

- Obtain the "raw" user communications requirements (UCR) data,
- Examine the "raw" UCR data for errors and describe the corrected UCR,
- Devise a simulation model which will develop communications network traffic data using the UCR,
- Develop the simulation model outputs into forms useful for guiding the communications system design, and
- Interpret the developed simulation model outputs.

A. "RAW" UCR ACQUISITION

The "raw" UCR were obtained by submitting appropriate forms (samples are shown on figures 1 and 2) and instructions to the agencies concerned with army support operations in a theater area.

The TACS Form I describes the radio and wire nets required for communications directly associated with intra-unit operations and, where necessary, the communications between adjacent echelons. Each distinct communication network in which subscribers in the unit participate has been entered on a separate TACS Form I. For each network, there are listed the number of subscribers, equipment allocated to each subscriber by quantity and type, type of installation, and a description of the network including purpose, communication range required, alternate means of communication, security requirements, recording capability, and conference capability.

TACS Form II develops a qualitative and quantitative description of the user communications requirements to be levied upon the common signal service provided by TACS-70. TACS Form II has been organized to show calling subscribers, facility or unit called, geographical distance between calling unit and called unit, specific subscriber in unit called, primary transmission mode (e.g., TP or TT), alternate transmission mode, equivalent call minutes for primary and alternate transmission mode, the traffic requiring security and priority handling, and appropriate comments to clarify specific communication requirements.

The UCR's, as organized, reflect individual user-to-user activity either on a unit-to-unit or discrete subscriber-to-subscriber basis. They describe basic communications needs unconstrained by any communications network considerations. For this reason, it is not necessary to first postulate a hypothetical communication "system," which must be tailored to accommodate the user requirements.

The UCR's are based, essentially, on experience in past and current operations, and, as necessary, extrapolation of these factors to mission oriented operational concepts planned for future implementation.

1. Communications Standards

The preparation of user requirements was based upon standards related to the operational environment of the theater units. The TACS standards serve primarily as a means of integrating unit mission and communications concepts. The standards consider both qualitative and quantitative descriptors of mission-communication integration. The data base construction, through application of these standards, should describe unit requirements in terms of type and quantity of each type of communications, its pattern involving the unit, and the identification of all traffic sources and sinks relative to the unit.

The TACS standards specify criteria for Direct Communication, Organic Communication, Conference Calls, Communication Range, Sole User Circuits, Security, Mobility, Hard Copy, Vulnerability to Electronic Countermeasures, Simplicity and Dependability, Alternate Facilities, and Command Post Fragmentation. Portions of the more important standards are described below.

- Direct Communication consists of either a sole-user channel or a net. The capability for direct communication will be limited to the next higher and lower echelons of command and of operational control.

- The capability to link users in conference calls or nets will be limited to those elements within a headquarters, or between units in a command, where there is a need for immediate coordination to control operations in progress.

- Sole user circuits can be justified only by an operational necessity for continuous contact between two stations.

- Requirements for secure communication must be justified by the operational importance of the traffic, the urgency of the messages, classification (considering both the sensitivity of the information and the length of time it will remain sensitive), and the volume and frequency of the traffic. The delay inherent in passing classified electrical messages, as well as the additional personnel and equipment required, warrants investigating the use of fast and frequent courier service as an alternate.

- Courier delivery will be relied upon as the primary means of transmission of hard copy. The use of electrical communications to produce hard copy at the terminating instrument is justified only when the urgency of the traffic precludes courier delivery and the generation of hard copy significantly reduces either the need for transcription, the probability of misinterpretation of important messages, or the time required to transmit and verify complex or lengthy messages.

2. Relating Different Transmission Modes

The UCR obtained in the survey describe the communications requirements for approximately 1000 units in terms of eight types of communications means including telephone, teletype, facsimile and data transmission facilities. Because of the multiple transmission modes, the analysis of both rate and volume loads becomes involved since each of the modes has its own quantitative parameters. A single parametric descriptor, equivalent call minutes (EQM), was used to simplify analysis. For all transmission modes specified, subscriber communications activity is expressed in EQM per 24 hours. For example for telephone, teletype and facsimile, the conversions to EQM are:

Telephone, in call minutes - 1 call minute	=	1 EQM
Teletype, in hundreds of words - 100 words	=	1 EQM
Facsimile, in page inches - 1 page inch	=	1 EQM

For traffic transmitted by messenger, the conversion factor used was one page inch of record traffic equals 1 EQM. This conversion simplified the task of equating traffic loads where traffic was diverted from electrical means to messenger.

It should be noted that only the channel requirements for traffic originating at the unit can be derived directly from the UCR's as submitted. To derive incoming traffic originated by other units, it was necessary to prepare the standard dominance or communications network matrix. For the typical theater army troop list, this matrix would be approximately 900 x 900. Manually developed subsets of this matrix are not feasible because of the large number of needlines that exist between units. An IBM 1620 was used to prepare summaries of incoming traffic. The sum of the incoming and outgoing traffic is tallied in terms of EQM to determine channel requirements.

B. CORRECTED UCR

It was necessary to evaluate the UCR's to determine if the unit requirements, as expressed, were minimal, but still adequate to support the unit missions, and to articulate the requirements in a way which permitted their translation into the requisite communications capability. Three methods of analysis were employed for review of the TACS UCR's:

- Comparison of the requirements stated for TACS units with those stated for similar units operating in a field army environment (TACOM 2/).

- Cost computation and comparative cost analyses of communications facilities.

- Quantitative analysis of communications flow.

1. TACS: TACOM Comparison

The relative comparison of communication requirements for like units (same TO&E) deployed in the theater and in the field army primarily served to examine UCR's as a function of mission and subordination. An analysis of all theater units was not possible since the troop lists for the theater and for the field armies differed considerably in unit types.

Trunk group requirements for similar units varied considerably depending upon the unit subordination. Possible reasons are: the unit requirements were prepared by different analysts, a period of a year or more occurred between preparation of TACOM and TACS requirements and missions and operations of units differ between the field army and the theater army. The preparation of requirements was based upon nearly identical standards, thus variations would not be expected as a result of differences in the criteria for specifying UCR.

2. Cost Evaluation

The cost computation and comparative analysis, involves preparation of four sets of cost data. These are the costs of organic communication facilities, cost ratios on a unit basis of organic costs per man (based on TO&E personnel strength), costs of the common signal service involving transmission facilities to support the unit, and cost ratios on a unit basis of common signal service per man. These cost figures and ratios were summarized on a service basis. By presenting this data by Service, i.e., Ordnance or Military Police, it was possible to examine relative communications costs assignable to particular operations within the theater areas.

The costs of the organic communications facilities for each unit were computed for the equipment listings (quantity and type) for each network specified for the unit. The costs of the common signal service support to the units were derived as a function of the number of calling subscribers, distance between calling and called subscribers, and minimum cost per channel kilometer for transmission facilities. The number of calling subscribers and the distance between calling and called subscribers were both based on theater troop lists and a specific deployment The costs per channel kilometer for transmission facilities used were developed as part of a parallel cost-effectiveness study 3/. The total costs of the common signal service support were divided by the unit personnel strength to obtain costs per man.

3. Communications Flow Evaluation

Although three methods were used to determine the validity of the "raw" UCR the most valuable one was to determine the average communications flow per man into and out of each unit in the theater area. The communications flow analysis provided quantitative data on the type and volume of each means of communication originated or received by using units.

Because of the large number of units and the transmission modes employed by each unit, an IBM 1620 was used to provide the tabulations required. The tabulations showed total equivalent traffic flow for each transmission mode (T_p, TT, etc.) in terms of equivalent call minutes (EQM). To provide an additional means of comparison of traffic flow, the ratio of total traffic flow to unit personnel strength was computed for each unit. These ratios were then grouped according to the following categories: less than 5 EQM/man, 5 to 10 EQM/man, 10 to 20 EQM/man, 20 to 30 EQM/man, 30 to 50 EQM/man, 50 to 100 EQM/man, and more than 100 EQM/man.

The magnitude of traffic loads for certain units, those appearing under the last category above, might require that each individual of the unit, on the average, be committed to hours of transmission time per day. The existence of such activity prompted special consideration of these units to determine if their UCR were valid. Usually the values of communications flow per man for different units could be compared, on a relative basis, to determine whether a given units communications demands were consistent with other similar units.

An evaluation of traffic flow data showed that various units in the theater area did have abnormally high traffic flows. Each of the agencies which had submitted UCR data leading to unusual traffic flows (unusual either in terms of its magnitude or its direction, e.g., a unit with moderate flow out of the unit but no traffic flow into the unit) participated in a review of the UCR to determine if the UCR had been properly prepared and submitted. The UCR review showed that certain UCR form instructions had not been correctly followed and the UCR of some units had been overstated. As a result of the UCR review corrected UCR were obtained which, in terms of total theater traffic, were reduced to 80% of the original UCR traffic requirements.

C. ALTERNATE MODE, PRECEDENCE AND SECURE TRAFFIC

In addition to the primary transmission requirements data which were obtained as a result of the UCR survey and data evaluation, data describing alternate transmission, precedence traffic and traffic/message security requirements were also evaluated.

1. Alternate Transmission Requirements

The user communications requirements, as stated, are predicated on particular transmission modes most suitable to support the operations of the user. To permit continuity of operations in the event of failure of the primary modes, alternate modes of transmission were specified. The alternate transmission mode requirements for each agency were tabulated to show the percentage of time traffic might be shifted to a particular secondary mode from a specific primary mode. These data were used to develop flow diagrams for each contributing agency describing the routing of primary to alternate traffic for all communications means. A composite flow diagram, summarizing the characteristics of all theater users is shown on figure 3.

2. Precedence Traffic

The User Communications Requirements submissions included estimates of types of traffic requiring precedence handling regarding: record traffic, voice (Tp) traffic, and operational communications. For each category the estimates have been stated as a percentage of total traffic. In general, for

record and voice (T_p) traffic, the major precedence requirements exist at major command headquarters and major operational units. Approximately 5% of the theater traffic requires precedence handling.

3. Secure Traffic

From a doctrinal standpoint, it might be desirable to provide cryptologic protection for all traffic regardless of the specific nature or content of the traffic. Since this objective cannot always be achieved because of technological, cost or operational limitations, users were requested to indicate traffic volumes (as a percentage of the total) which would specifically require cryptologic protection. Approximately 3% of the total theater traffic required cryptographic protection.

D. SIMULATION MODEL DESIGN

The computer simulation model accepts, as input data, the UCR, the locations of the theater units, and a description of the communications network. The network is described in terms of network nodes, (major network transmission centers) the node locations and the sets of node connections (see figure 4). The program will accept up to 1200 units, networks containing 40 nodes with 6 connections per node and unit and node location input data given in Universal-transverse-mercater (UTM), military-reference-grid-system (MRGS) coordinates.

The simulation model routes traffic by first assigning each unit to its nearest network node. Traffic is then routed from communications source unit to destination units so that it passes through a minimal number of intermediate nodes. The total traffic, by type (telephone, data, etc.), passing over each network link as a result of satisfying all the UCR is compiled and obtained as the primary program output.

In order to determine the distribution of traffic over the network the program only compiles that traffic which has passed over N or fewer links. Thus, by performing program runs with different values of N the distribution of local and long distance traffic can be determined for any desired network for each communications means (telephone, teletype, facsimile, etc., see figure 5).

E. DEVELOPING TRAFFIC RELATIONSHIPS USING THE SIMULATION MODEL

The simulation model supplies, directly, the data describing the traffic loading on each network node and link for each communication means. By using the direct program outputs obtained for various network configurations, unit and network deployments and network modifications, the following data can be developed.

- Communications survivability for various levels of network destruction (figure 6).

- Relative switching loads for different networks for various levels of network destruction (figure 7).

- Network efficiency and its sensitivity to network and deployment perturbations (figure 8).

- The distribution of link channel requirements can be determined for different networks. This data can be used to describe the efficiency of equipment utilization in specified networks (figure 9).

III. RESULTS

The simulation model was used to determine communications traffic characteristics for a variety of networks subjected to the survey determined communications data. Specifically the model was used to compare seven different theater communications networks consisting, nominally, of thirty-five network nodes with an average of three network connections per node. The model results showed:

- Which of five networks developed from cost-effectiveness evaluations of available communications equipments best supported the surveyed theater communications traffic requirements.

- The total theater traffic tends to be uniformly distributed over the theater area. This characteristic is shown by the uniform slope of the curves (figure 10) describing the distribution of traffic over specific numbers of network links.

- The communications networks examined are relatively insensitive to changes in network location. This characteristic is shown (figure 10) by the lack of significant changes in traffic load for each of the deployment situations studied. The four deployment situations studied (A,B,C,D) represent displacing the communications networks 15, 30 and 45 km from a reference network location. Thus for each network displacement the complex of theater units homing on particular network nodes was changed. The relative insensitivity to the network deployment changes is primarily a result of the relatively uniform distribution of theater traffic.

- The bulk (60.5%) of the theater traffic is carried by telephone communication. The distribution of theater traffic by communications type is shown on figure 11. The individual traffic

distributions for each communications means also tend to be uniformly distributed. There are, however, some generalizations which can be made about the relative long-local traffic distribution of the various communications means.

- Telephone, facsimile and data traffic all have similar long distance traffic behavior.
- Data traffic has relatively little local traffic loading.
- Teletype and messenger traffic tend to have a greater portion of long distance traffic.

A sample of network trunk requirements determined using the simulation model is shown in figure 12. The distributions of required trunk loads for the communications specified in the UCR are shown in figure 13.

The above analysis led to the design of a theater communication system using surveyed subscriber data, where the distribution and characteristics (quantities of secure and priority traffic, alternate message routing) of traffic could be determined and used to measure the likelihood that a given network would satisfy the communications requirements of the theater subscribers.

ACKNOWLEDGMENTS

The helpful support of Lt. Col. O'dell of the Combat Service Support Group (USACDC) in the review of the UCR, and of Messrs. L. Northrop and C. Raphun of the communications electronics agency (Fort Huachuca, Arizona) in the evaluation of the system designs was critical to the successful completion of the TACS program.

The contributions of Dr. R. Crawford, Dr. M. Brilliant and Mr. R. Quenstedt were especially important in developing the system design techniques.

REFERENCES

1/ Combat Developments Study Directive: "Theater Army Communication System Requirements, 1965-1970 (TACS-70) (U)," U.S. Army Combat Developments Command, Fort Belvoir, Virginia, 16 February 1966.

2/ "Field Army Requirements for Tactical Communication (TACOM) (U)," November 1964. Final Draft, U.S. Army Combat Developments Command, Fort Belvoir, Virginia.

3/ R.L. Crawford, "Cost-Effectiveness Techniques for a Theater Army Communications System," March 1966, 1966 U.S. Army Operations Research Symposium, Fort Monmouth, New Jersey.

TACS REPORTING FORM I

ORGANIC COMMUNICATIONS

UNIT Transportation Diesel-Electric Locomotive Repair Company   PROPONENT Transportation Agency

Net Number and Designation 55-237 A and B

Purpose: Command and control within the unit and wire communication with higher headquarters and supported units

Range: N/A   Alternate Means: Messenger & radio

Remarks:

STA NR	USER	EQUIPMENT QTY	NOMEN	INST	REMARKS	
	NET ORGANIZATION					
1.	Company Commander	1	TA-312/PT	1	User	4862
2.	Diesel Loc SH Supt.	1	"	"	"	4862

NET DESCRIPTION

Concept.

a. Purpose. To provide wire communications within the unit and to higher headquarters and supported units.

b. Range. NA.

c. Alternate Means. Messenger and radio.

d. Security Requirements. None.

e. Recording Capability. None required.

f. Conference Capability. None require.

Figure 1

TACS Reporting Form I

TACS REPORTING FORM II
DEMANDS ON THE COMMON SIGNAL SERVICE

PROPONENT: Transportation Agency

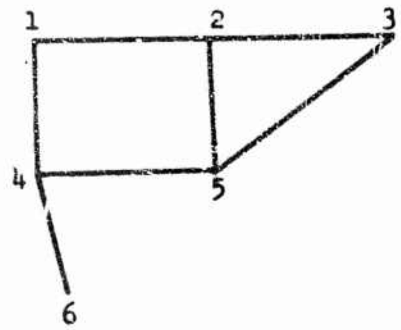
UNIT: Transportation D-E Loccomotive Repair Company

PAR NR	SUBSCRIBER	FACILITY CALLED	DIST	SUBSCRIBER CALLED	LINE NR	DES MODE	QTY	ALT MODE
1.	Co. Commander	Bn Hq	100	Bn Adjutant (5)	1	Tp	5	Radio
			100	Bn Commander(5)	2	Tp	5	Radio
			100	Bn S-4 (5)	3	Tp	5	Radio
			100	Bn S-3 (5)	4	Tp	5	Radio
			25	Supply Co. (5)	5	Tp	5	Radio
2.	Diesel Loc Sh Supt		100	Bn S-4 (5)	1	Tp	6	Radio
			100	Bn CO (5)	2	Tp	5	Radio
			50	Co (4)	3	Tp	10	Radio

NOTE: It is most important that this unit be able to communicate with all units that they support over the 300 miles of railroad that is in their support area. Lack of communication will contribute to an unsatisfactory maintenance posture.

Figure 2

TACS Reporting Form II



NETWORK

NETWORK DESCRIPTION

Node	Node Connections	Node Location
1	2,4	31VTU0173
2	1,3,5	31VTU4779
3	2,5	.
4	1,5,6	.
5	2,3,4	UTM-MRGS
6	4	Coordinates

Figure 4

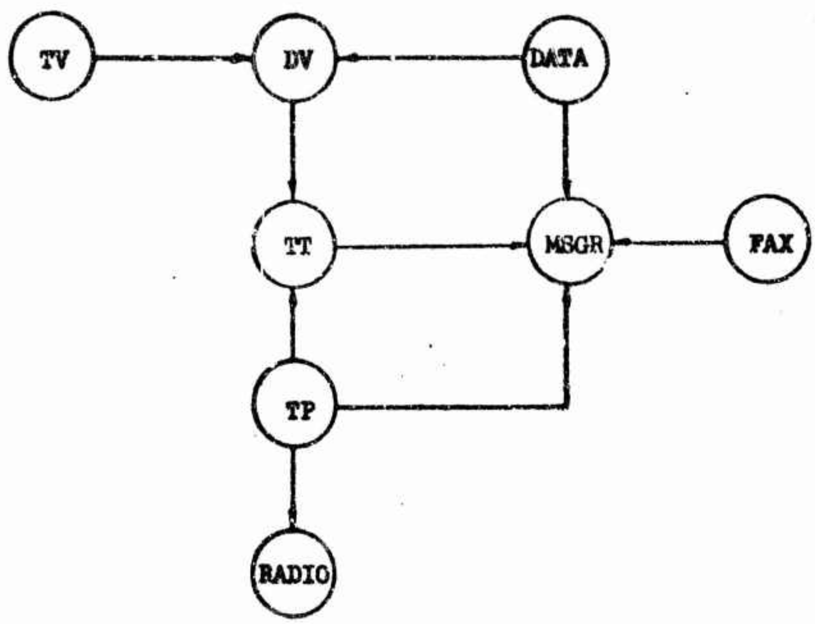


Figure 3

Summary

User Preference, Choice of Alternate Transmission Modes

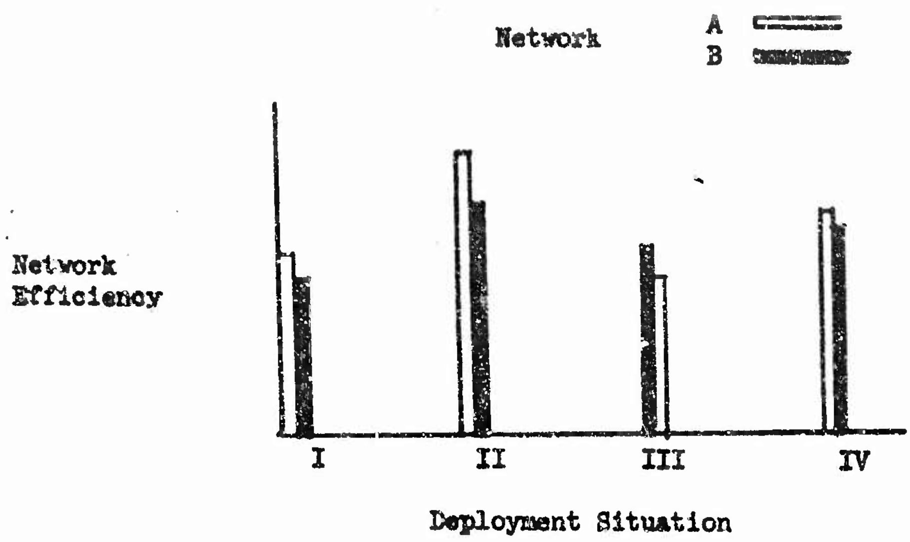


Figure 5

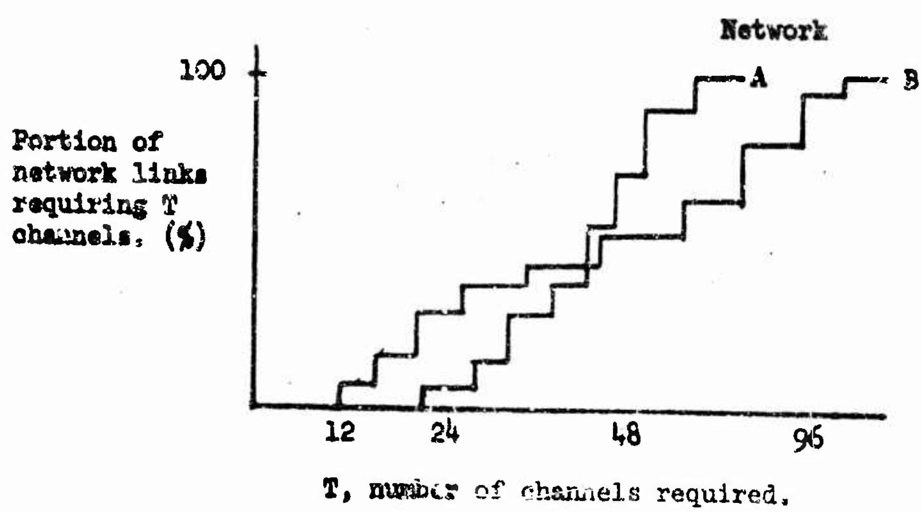


Figure 6

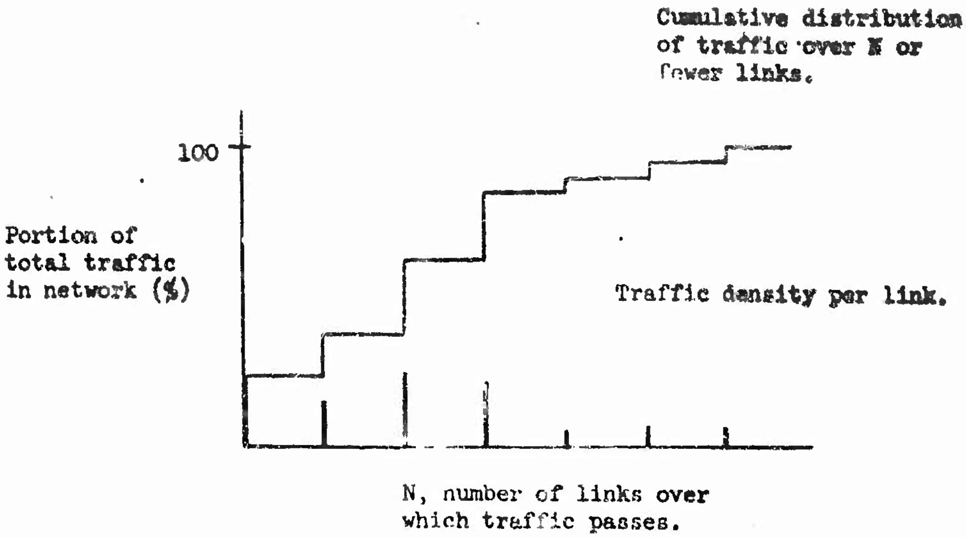


Figure 7

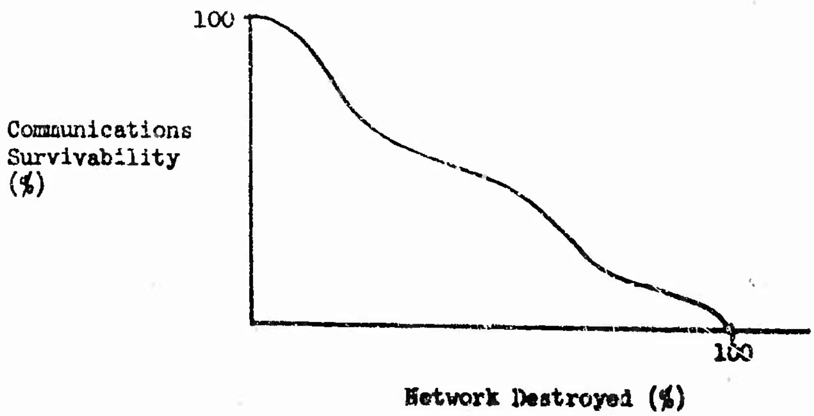


Figure 8

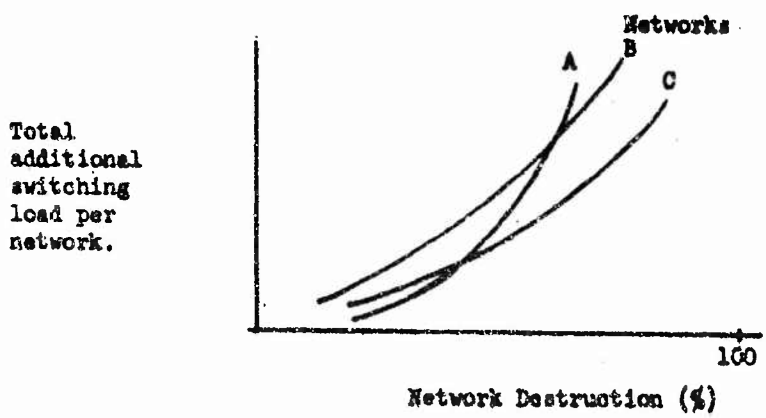
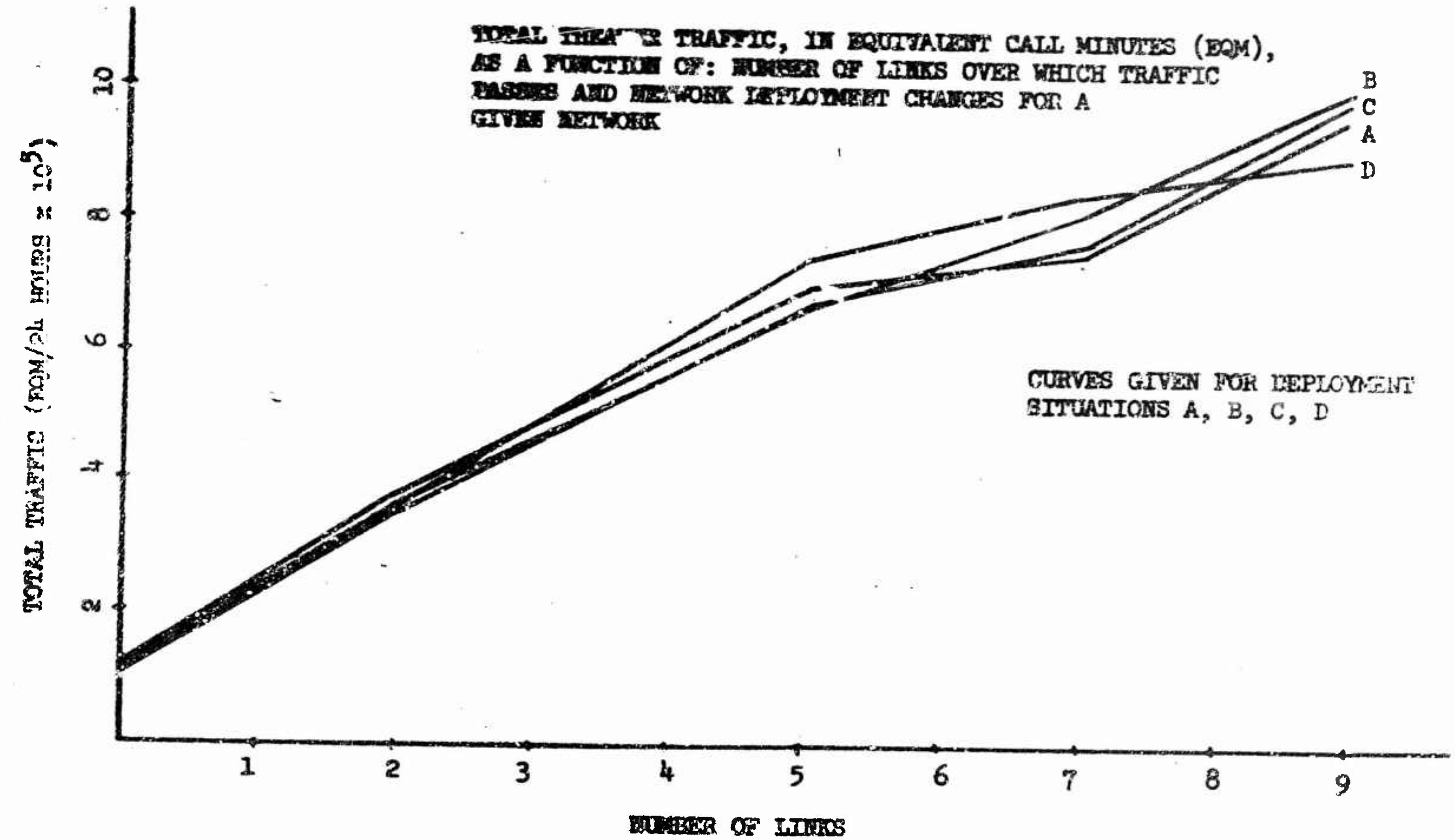
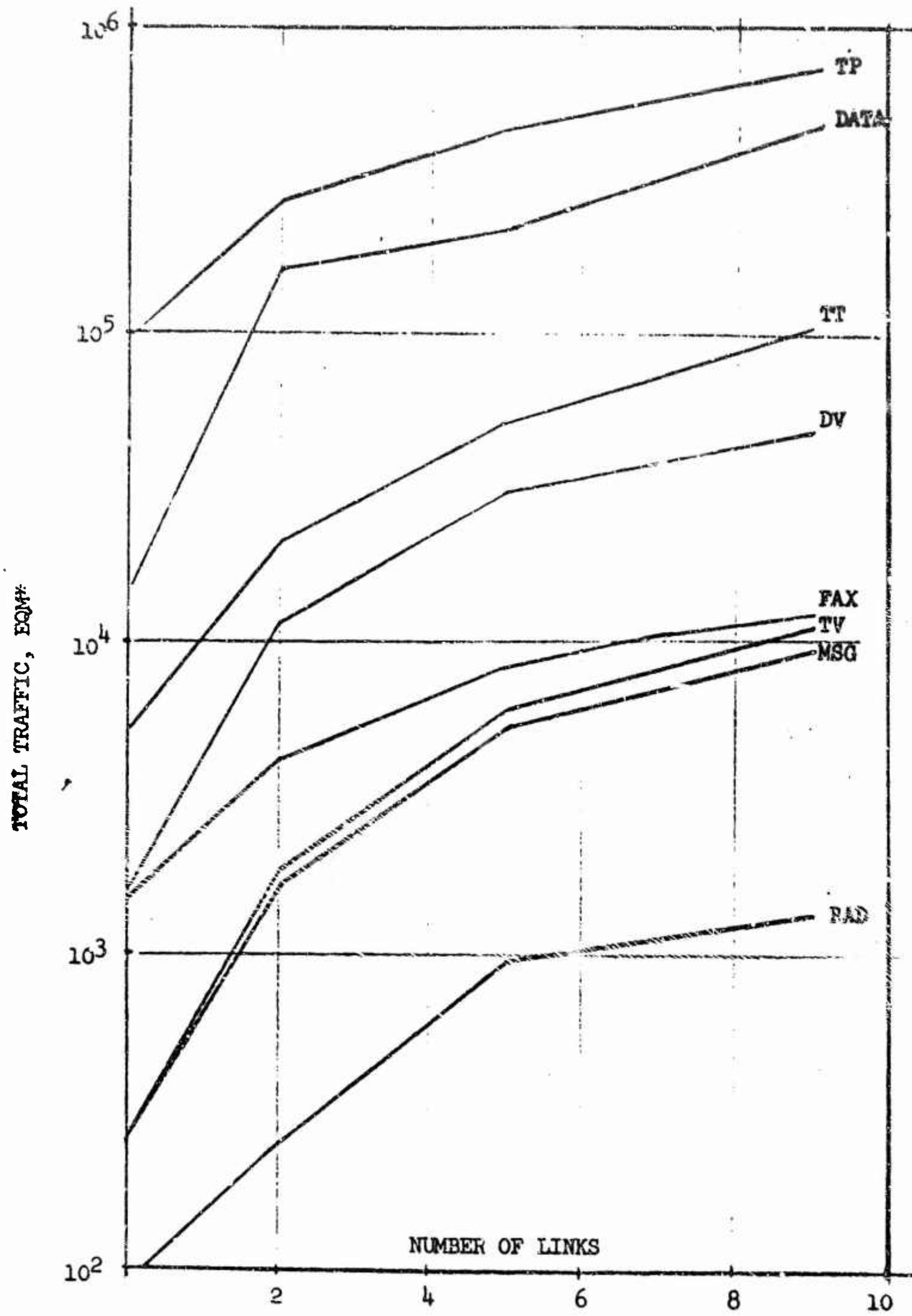


Figure 9

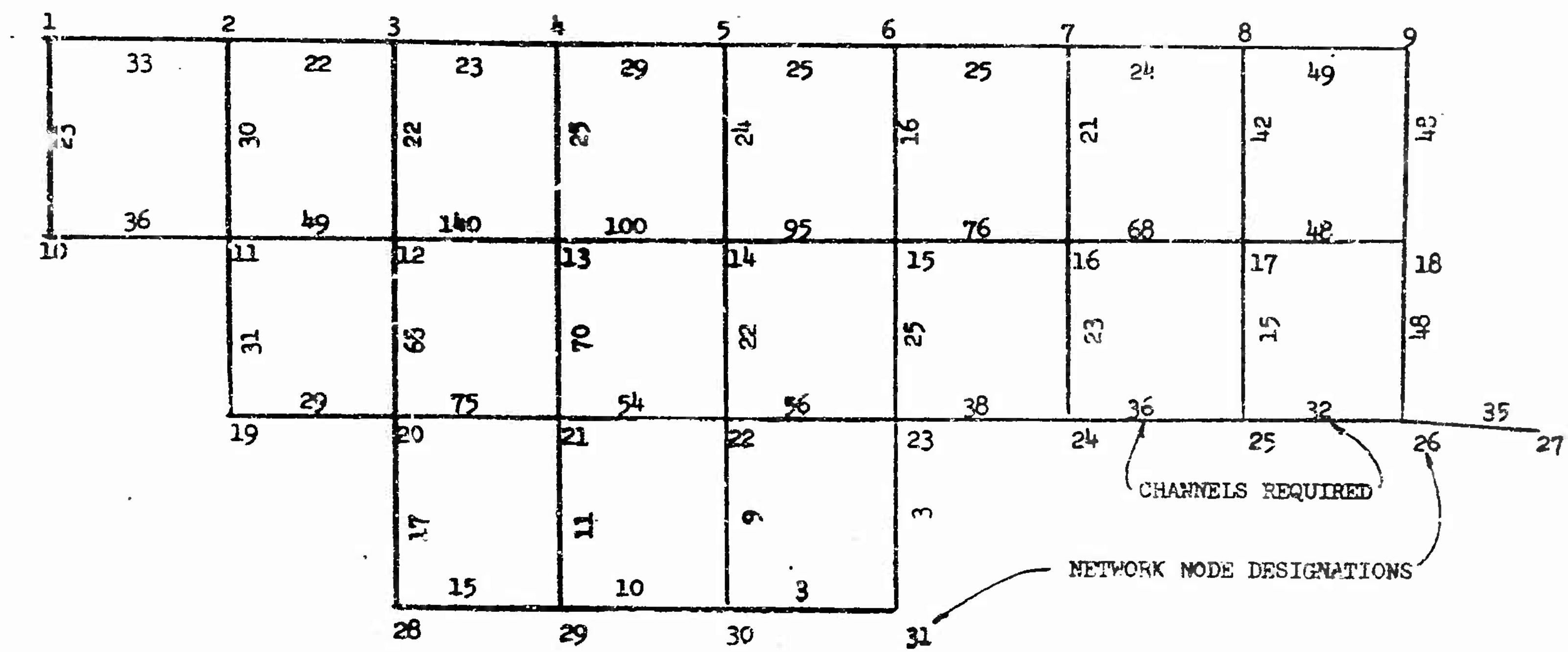

TOTAL THEATER TRAFFIC, IN EQUIVALENT CALL MINUTES (EQM), AS A FUNCTION OF: NUMBER OF LINKS OVER WHICH TRAFFIC PASSES AND NETWORK DEPLOYMENT CHANGES FOR A GIVEN NETWORK
CURVES GIVEN FOR DEPLOYMENT SITUATIONS A, B, C, D
TOTAL TRAFFIC (EQM/24 HOURS x 10^5)
NUMBER OF LINKS
B
C
A
D
2
4
6
8
10
1
2
3
4
5
6
7
8
9



TOTAL TRAFFIC, IN EQUIVALENT CALL MINUTES*, FOR EACH COMMUNICATION MEANS AS A FUNCTION OF THE NUMBER OF NETWORK LINKS OVER WHICH TRAFFIC IS PASSED

* EXCEPTION - DATA TRAFFIC IS GIVEN IN KILOBITS PER 24-HOUR PERIOD

FIGURE 11



NUMB'R OF CHANNELS REQUIRED PER NETWORK LINK TO SATISFY THE UCR WITH A P-3 GRADE OF SERVICE

FIGURE 12

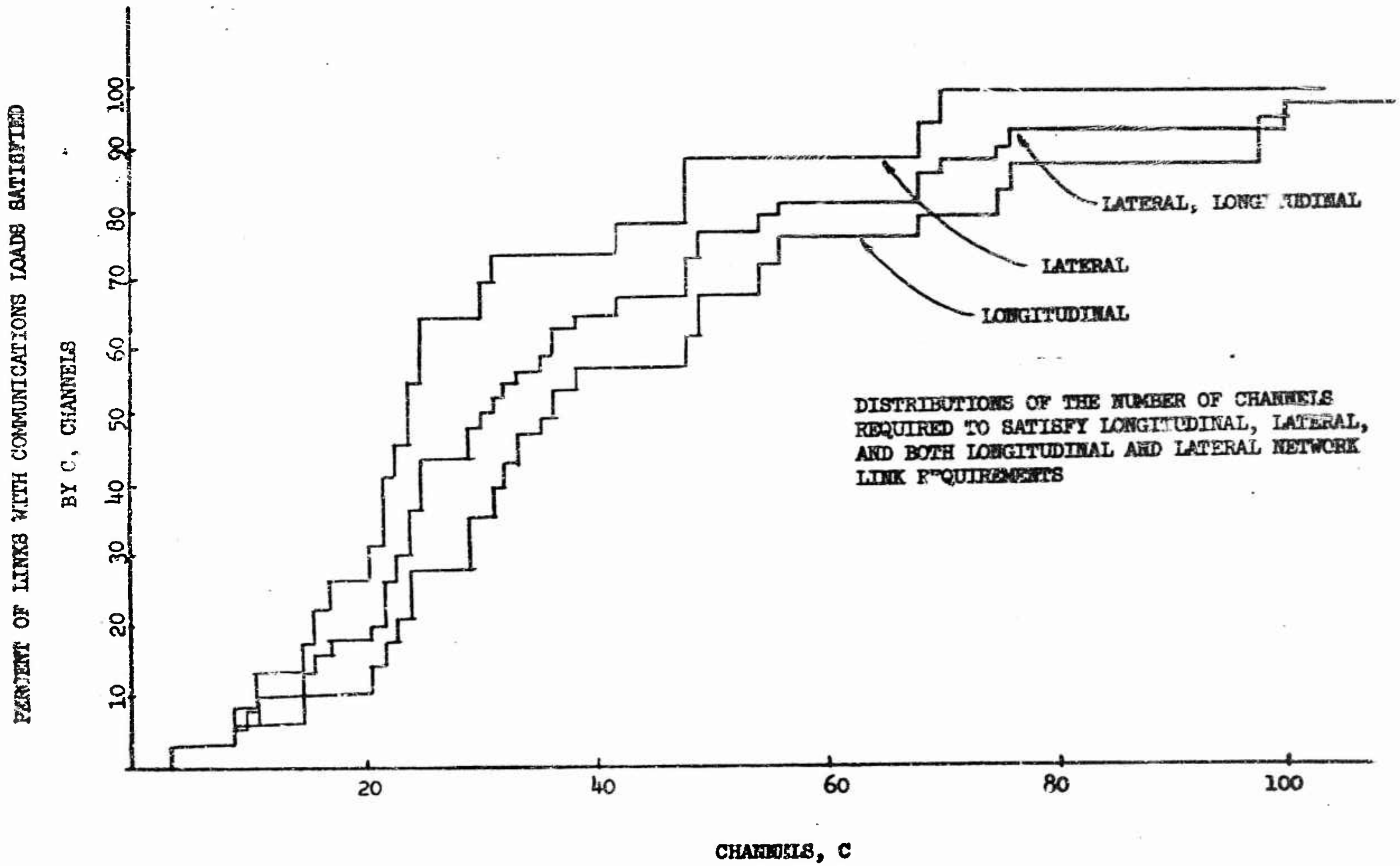

PERCENT OF LINKS WITH COMMUNICATIONS LOADS SATISFIED BY C, CHANNELS
10
20
30
40
50
60
70
80
90
100
20
40
60
80
100
CHANNELS, C
LATERAL, LONGITUDINAL
LATERAL
LONGITUDINAL
DISTRIBUTIONS OF THE NUMBER OF CHANNELS REQUIRED TO SATISFY LONGITUDINAL, LATERAL, AND BOTH LONGITUDINAL AND LATERAL NETWORK LINK REQUIREMENTS

FIGURE 13

INCLINATION AND NODAL CROSSING OF THE ORBITAL PLANE FROM RADAR DATA

by

P. J. SHORT, RISO

and

W. L. SHEPHERD, ID

January 1966

US ARMY TEST AND EVALUATION COMMAND

RANGE INSTRUMENTATION SYSTEMS OFFICE
WHITE SANDS MISSILE RANGE
NEW MEXICO

ABSTRACT

The paper utilizes a well-known theorem concerning minima for a certain class of quadratic forms of three variables with a normalizing constraint to determine the angle of inclination and the right ascension of nodal crossing of the orbital plane of a near earth satellite from observation data.

INTRODUCTION

It is a well-known fact that the path of motion of a near earth satellite is an ellipse with the center of the earth at one focus. Hence, the path of motion is contained in a plane which passes through the center of the earth.

The orbit is determined by the following six parameters (orbital parameters):

1. Inclination (i) of the orbital plane to the earth's equatorial plane
2. Right ascension of nodal crossing (R_Ω)
3. Period (T)
4. Semi-major axis (a)
5. Eccentricity (e)
6. Argument of perigee (ω).

The last three quantities determine the orientation and shape of the orbit in the orbital plane while the first two determine the orientation of the orbital plane with respect to the earth's equatorial plane. The third quantity is the time required for the satellite to complete one revolution.

The purpose of this paper is to describe a simple technique for computing i and R_Ω from observation data. The technique is then illustrated by examples using radar data which were recorded at White Sands Missile Range, New Mexico.

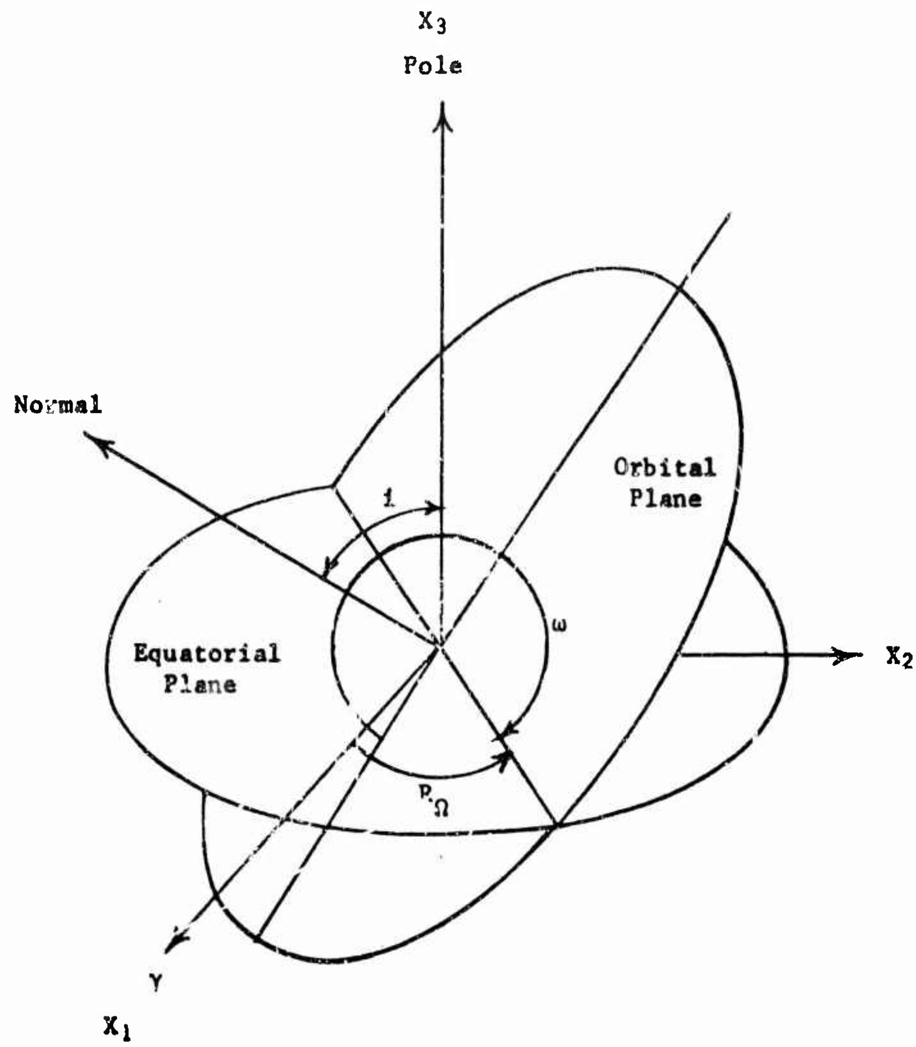


FIGURE 1 INERTIAL COORDINATE SYSTEM

The plane of orbital motion will be determined with respect to an inertial coordinate system (Figure 1) with its origin at the center of the earth. In this system the coordinate axes are defined as follows:

1. The positive X_1 axis points in the direction of the Vernal Equinox γ.

2. The positive X_3 axis points in the direction of the North Pole P.

3. The positive X_2 axis is chosen so that the system is right handed.

With this notation the path of motion is contained in a plane defined by an equation of the form:

$$\sum_{j=1}^{3} \alpha_j X_j = 0 \tag{1}$$

where α_1, α_2, and α_3 are constants. In fact, it is easy to show that

$$\begin{aligned} \alpha_1 &= K \sin R_\Omega \sin i \\ \alpha_2 &= K \cos R_\Omega \sin i \\ \alpha_3 &= K \cos i \end{aligned} \tag{2}$$

where K is a constant which we may assume to be unity. Hence, if α_1, α_2, and α_3 are known, it is easy to find i and R_Ω. Thus, the problem is reduced to finding α_1, α_2, and α_3 from observation data.

The method used in this report is based on the Method of Least Squares and a well-known theorem which can be found in almost any advanced calculus text. A theorem concerning minima for a certain class of quadratic forms of three variables with a normalizing restraint is proved and utilized in an artificial earth satellite problem.

Using the convention that R^n denotes Euclidean n-space and if $x \epsilon R^n$ then $x = (x_1, x_2, \ldots, x_n)$ where x_j ($j = 1, 2, \ldots, n$) is the jth coordinate of the point x. If there are several, say M, points under consideration, then a superscript will be used. Hence, the mth point of a set of M points will be represented by

$$x^m = (x_1^m, x_2^m, \ldots, x_n^m)$$

where $m = 1, 2, \ldots, M$.

If A is a matrix then A* is its transpose.

THE MINIMUM PROBLEM

Let $S = \{\beta \in R^3 \mid g(\beta) = 1\}$ and let $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ with $\alpha \in S$. Define

$$g(\alpha) = \alpha\alpha^* \tag{3}$$

and

$$f(\alpha) = \alpha B \alpha^* \tag{4}$$

where B is a real symmetric matrix. $\min_{\alpha \in S} f(\alpha)$ exists (Reference 1, pages 196-198), and Lagrange's method applies.

Lagrange's method is used to prove

THEOREM 1: $\min_{\alpha \in S} f(\alpha)$ is the least characteristic root of B and is attained at the corresponding normalized characteristic vector of B for which $\alpha_3 \geq 0$.

PROOF: Notice that symmetry of f and g indicates that $\alpha_3 \geq 0$ is a permissible restriction. Lagrange's function is

$$F(\alpha, t) = f(\alpha) - t g(\alpha), \tag{5}$$

and Lagrange's equations are

$$g(\alpha) = 1, \tag{6}$$

$$\nabla F(\alpha, t) = \nabla(\alpha B \alpha^* - t\alpha\alpha^*) = 0. \tag{7}$$

Hence from Equation (7)

$$B\alpha - t\alpha = 0 \tag{8}$$

Let t_j and v_j $(j = 1, 2, 3)$ denote the characteristic roots of B and the corresponding normalized vectors of B for which $\alpha_3 \geq 0$. The stationary values of $f(\alpha)$, $\alpha \in S$, with $\alpha_3 \geq 0$ occur in the set $f(v_1)$, $f(v_2)$, $f(v_3)$. Hence (Ref 2 P220)

$$\min_{\alpha \in S} f(\alpha) = \min_{j=1,2,3} f(v_j) \tag{9}$$

Note that $F(\alpha,t)$ is homogeneous of degree two in α_1, α_2, and α_3. Eulers' theorem is applicable here and yields

$$1/2\ (\nabla F)\ .\alpha = F$$

then

$$1/2\ \nabla F\ (v_j,\ t_j)\ .\ v_j = 0 \qquad (10)$$

or

$$f\ (v_j) - t_j g\ (v_j) = 0. \qquad (11)$$

Since $g(v_j) = 1$, we have

$$t_j = f(v_j) \qquad (12)$$

and

$$\min_{j=1,2,3} f(v_j) = \min_{j=1,2,3} t_j \qquad (13)$$

Equations (9) and (13) yield the theorem.*

* The generalization to R^n is obvious.

THE ORBIT PROBLEM

Let x^k (k=1, ..., N) be N points or positions of a near earth satellite, referred to the inertial coordinate system described in the Introduction. Now find the least squares fit of this set of points to the plane defined by Equation (1) with the restraint

$$\alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 1, \quad \alpha_3 > 0 \tag{14}$$

Substitution of x_j^k into Equation (1) yields

$$\sum_{j=1}^{3} \alpha_j \, x_j^k \quad 0; \; k = 1, 2, \ldots, N \tag{15}$$

The Method of Least Squares requires the α for which

$$f(\alpha) = \sum_{k=1}^{N} \left(\sum_{j=1}^{3} \alpha_j \, x_j^k \right)^2 \tag{16}$$

is a minimum subject to the condition given by Equation (14). Thus

$$f(\alpha) = \sum_{m,j=1}^{3} b_{mj} \, \alpha_m \, \alpha_j \tag{17}$$

where

$$b_{mj} = \sum_{k=1}^{N} x_m^k \, x_j^k \tag{18}$$

Now, $B = (b_{mj})$ is a real symmetric matrix and $f(\alpha) = \alpha B \alpha^*$. Hence, the preceding theorem applies directly.

THE MINIMUM CHARACTERISTIC ROOT

In order to complete the problem the minimum root of Equation (8) must be found. Equation (8) can be written as:

$$b_{mj} - \delta_{mj} t = 0, \tag{19}$$

where δ_{mj} is Kroneckers' delta. Equation (19) has a non-trivial solution if and only if

$$\det (b_{mj} - \delta_{mj} t) = 0. \tag{20}$$

Clearly, Equation (20) yields a cubic in t,

$$t^3 + C_2 t^2 + C_1 t + C_o = 0, \tag{21}$$

the roots of which are real. The coefficients C_n (n=0,1,2) are functions of the b_{mj}.

THEOREM 2: The minimum root of the cubic, Equation (21), is given by

$$t_m = \frac{-C_2}{3} - 2\sqrt{\frac{|A|}{3}} \cos \left[\frac{2\phi - (1 + (-1)^{m+1})}{6}\right]\Pi \tag{22}$$

where

$$A = C_1 - \frac{C_2^2}{3} \tag{23}$$

$$B = C_o - \frac{C_1 C_2}{3} + \frac{2C_2^3}{27} \tag{24}$$

$$m = \frac{1}{2}\left(1 + \frac{B}{|B|}\right), \text{ i.e., } m = 0 \text{ or } 1, \tag{25}$$

and ϕ is the first quadrant angle determined by

$$\cos = \sqrt{\frac{27B^2}{4|A|^3}}$$

PROOF: To prove the theorem, use Cardan's solution of the cubic. Thus, setting $t = s - C_2/3$ reduces Equation (21) to

$$s^3 + As + B = 0 \tag{26}$$

where A and B are defined by Equations (23) and (24). Then since the roots of Equation (21) are real

$$\Delta = \frac{B^2}{4} + \frac{A^3}{27} \leq 0 \tag{27}$$

which implies A<0.

Now there are two cases to consider separately.

If $\Delta<0$, then the roots of Equation (26) are $(p = 2\sqrt{|A|/3})$ (28)

clearly, $s_k = (-1)^m \, p \cos\left(\frac{\phi + 2k\Pi}{3}\right)$, $k = 0,1,2.$

$m=0$ implies $B<0$. Hence, $s_1<s_0$ and $s_1<s_2$ (29)

or

$m=1$ implies $B>0$. Hence, $s_0<s_1$ and $s_0<s_2$ (30)

so that either s_0 or s_1 is the minimum root.

But,

$$s_1 = p \cos\left(\frac{\phi}{3} + \frac{2\Pi}{3}\right) = p \cos\left(\frac{\phi-\Pi}{3} + \Pi\right) = -p \cos\left(\frac{\phi-\Pi}{3}\right) \tag{31}$$

or

$$s_1 = p \cos\left(\frac{2\phi - 2\Pi}{6}\right) = -p \cos\left\{\frac{2\phi - [1+(-1)^2]\Pi}{6}\right\}, \tag{32}$$

and

$$s_0 = -p \cos\frac{\phi}{3} = -p \cos\left\{\frac{2\phi - [1+(-1)]\Pi}{6}\right\} \tag{33}$$

Hence,

$$s_m = -p \cos\left\{\frac{2\phi - [1+(-1)^{m+1}]\Pi}{6}\right\} \tag{34}$$

If $\Delta = 0$, then $\cos\phi = 1$ or $\phi = 0$. Thus, the roots of the cubic are

$$s_k = (-1)^m p \cos\frac{2\Pi k}{3}, \; k = 0,1,2 \tag{35}$$

clearly,

$m=0$ implies $B<0$. Hence, $s_0>s_1 = s_2$ (36)

or

$m=1$ implies $B>0$. Hence $s_o < s_1 = s_2$ (37)

In either case it is easy to see that

$$s_m = -p \cos \left[\frac{(1+(-1)^{m+1}) \Pi}{6} \right] \quad (m=0,1), \tag{38}$$

This completes the proof.

SUMMARY

It is to be understood that all observational data is expressed in an inertial coordinate system.

It is required to obtain approximations to the inclination (i) and right ascension of nodal crossing (R_Ω) defining the orientation of the orbital plane with respect to the earth's equatorial plane from a set $\left\{(x_1^j, x_2^j, x_3^j) \mid j=1,2, \ldots, n\right\}$ of observations of a near earth satellite.

The physical constraint on the satellite is that its orbit must be contained in a plane defined by the equation:

$$\sum_{k=1}^{3} \alpha_k \; x_k^j = 0 \tag{39}$$

where

$$\left\{\begin{aligned} \alpha_1 &= \sin R_\Omega \sin i \\ \alpha_2 &= \cos R_\Omega \sin i \\ \alpha_3 &= \cos i \end{aligned}\right. \tag{40}$$

In addition, the constraints

$$\alpha_1^2 + \alpha_2^2 + \alpha_3^2 - 1 = 0 \text{ and } \alpha_3 \geq 0 \tag{41}$$

are required.

Define

$$f(\alpha) = \sum_{k=1}^{N} \left(\sum_{j=1}^{3} \alpha_j x_j^k \right)^2 \tag{42}$$

$$h(\alpha) = \sum_{j=1}^{3} \alpha_j^2 \tag{43}$$

$$g(\alpha) = h(\alpha) - 1 = 0 \tag{44}$$

The α_j are obtained by means of the Method of Least Squares and Lagrangian multipliers, i.e.,

$$\nabla f - t \nabla g = 0 \tag{45}$$

or

$$\frac{\partial f(\alpha)}{\partial \alpha_k} - t \frac{\partial h(\alpha)}{\partial \alpha_k} = 0 \text{ , } k = 1,2,3 \tag{46}$$

define

$$b_{mj} = \sum_{k=1}^{N} x_m^k x_j^k \tag{47}$$

then from Equation (46)

$$\begin{pmatrix} b_{11} - t & b_{12} & b_{13} \\ b_{12} & b_{22} - t & b_{23} \\ b_{13} & b_{23} & b_{33} - t \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = 0 \tag{48}$$

has non trivial solutions if and only if

$$\begin{vmatrix} b_{11} - t & b_{12} & b_{13} \\ b_{12} & b_{22} - t & b_{23} \\ b_{13} & b_{23} & b_{33} - t \end{vmatrix} = 0 \tag{49}$$

The smallest root of this equation is given by

$$t_m = \frac{C_2}{3} - 2\sqrt{\frac{|A|}{3}} \quad \cos\left[\frac{2\phi - (1+(-1)^{m+1})}{6}\,\pi\right] \tag{50}$$

where

$$A = C_1 - \frac{C_2^2}{3} \tag{51}$$

$$B = C_0 - \frac{C_1 C_3}{3} + \frac{C_2^3}{27} \tag{52}$$

$$m = \frac{1}{2}\left(1 + \frac{B}{|B|}\right) \tag{53}$$

and ϕ is the acute (positive) angle determined by the equation

$$\cos\phi = \sqrt{\frac{27\ B^2}{4|A|^3}}$$

The problem was programmed in Fortran Language for the IBM 1620 computer. In spite of the formulas given in the preceding sections for computing the minimum characteristic root, an iterative technique was found more convenient and accurate to use.

The technique, a simple application of the Principle of Contraction Mappings, is illustrated in Reference 3 and proved in Reference 4. The original equations are written as follows:

$$t^1 = (b_{13}\ \alpha_1^1 + b_{23}\ \alpha_2^1 + b_{33}\ \alpha_3^1)/b_{33} \qquad (54)$$

$$\alpha_1^1 = (b_{11}\ \alpha_1^1 + b_{12}\ \alpha_2^1 + b_{13}\ \alpha_3^1)/t^1 \qquad (55)$$

$$\alpha_2^1 = (b_{12}\ \alpha_1^1 + b_{22}\ \alpha_2^1 + b_{23}\ \alpha_3^1)/t^1 \qquad (56)$$

To find the fixed point, arbitrary values are assigned to α_1^1 and α_2^1 and fix α_3^1 (=1), and the above set of equations is used.

$$(\alpha_1^1,\ \alpha_2^1,\ \alpha_3^1,\ t^1),\ \alpha_3^1 = 1.$$

Next, if $(\alpha_1^2, \alpha_2^2, \alpha_3^2, t^2)$ is a second solution, $t^1 \neq t^2$, then α_3^2 is eliminated from the original group of equations using the fact that

$$\alpha_1^1\ \alpha_1^2 + \alpha_2^1\ \alpha_2^2 + \alpha_3^1\ \alpha_3^2 = 0 \qquad (57)$$

The following set of equations

$$\alpha_1^1 \ \alpha_1^2 + \alpha_2^1 \alpha_2^2 + \alpha_3^1 \ \alpha_3^2 = 0 \tag{58}$$

$$t^2 = (b_{12} \alpha_1^2 + b_{22} \alpha_2^2 + b_{23} \alpha_3^2)/\alpha_2^2 \tag{59}$$

$$\alpha_1^2 = (b_{11} \alpha_1^2 + b_{12} \alpha_2^2 + b_{13} \alpha_3^2)/t^2 \tag{60}$$

may be used to find $(\alpha_1^2, \alpha_2^2, \alpha_3^2, t^2)$. Fix α_2^2 (=1) and let α_1^2 be arbitrary and find the fixed point of the above set of equations. Continue in a similar manner to find $(\alpha_1^3, \alpha_2^3, \alpha_3^3, t^3)$.

Let c_1, c_2, and c_3 be arbitrary constants, then $(c_1 \alpha_1^1, c_1 \alpha_2^1, c_1 \alpha_3^1, t^1)$, $(c_2 \alpha_1^2, c_2 \alpha_2^2, c_2 \alpha_3^2, t^2)$, and $(c_3 \alpha_1^3, c_3 \alpha_2^3, c_3 \alpha_3^3, t^3)$ represent all possible solutions to the system of equations, and clearly c_j ($j = 1,2,3$) is determined by the conditions that

$$c_j^2 \left[(\alpha_1^j)^2 + (\alpha_2^j)^2 + (\alpha_3^j)^2 \right] = 1 \tag{61}$$

and

$$\alpha_3^j \geq 0. \tag{62}$$

From Reference 4, the values for j=3 are those which will make f(α) a minimum.

Table I shows results when computed from actual sampling rates, and Table II shows results when computed using one point per six seconds. All data were recorded at WSMR using AN/FPS-16 radars.

The column labeled pass in the tables is not the actual pass over WSMR. All trackings were ordered according to date and time and then numbered. Thus, Pass Number 4 might actually have been Pass Number 10. Omitted numbers in this column mean that the data for that particular pass were not usable.

In some cases an A is placed after the pass number. This indicates that these data were uncorrected for refraction, etc.

An asterisk in the pass column signifies there are known errors (bad data) in the data which have not been removed.

The results could be improved upon if more accurate procedures had been used; i.e., multiple precision. This is apparent in comparing t and f(t) which should agree and should be non-negative.

In general, even with bad data present and loss of accuracy in computing, the results are not too bad.

Unfortunately, no numbers for i and R_Ω were available for comparison.

TABLE I

Satellite	Pass	No. of Points	i(degrees)	R_Ω(degrees)	$10^4 t$	$10^4 f(t)$	Radar
TR-4B	1*	4358	32.123	190.466	2960.000	2950.000	R-113
	1	1610	32.585	188.527	1220.000	1208.000	R-114
MA-5	2	464	32.560	153.790	1.700	0.112	
MA-6	2	192	40.525	264.460	2.010	1.882	
MA-7	1*	1505	33.277	297.484	195.000	194.800	R-113
	2	1373	32.547	294.555	0.000	-0.478	
	2	389	32.539	294.390	0.100	0.071	R-114
	3	1292	32.605	294.552	2.200	1.921	R-113
SA-5	1	1294	31.399	170.755	6.700	1.050	
	3*	1213	30.912	150.700	89.800	89.120	
	4	1088	31.247	139.750	15.800	14.490	
	5	575	31.107	135.320	5.100	5.146	
	6	1985	31.493	133.182	19.000	18.720	
	7	216	32.202	45.191	0.200	0.200	
	8	175	76.931	114.317	0.750	0.732	
	9	630	31.005	26.674	1.800	1.978	
	10	3159	31.444	28.696	86.000	89.100	
	11	2990	31.448	28.373	122.000	124.200	
	13	2189	31.685	341.725	107.500	107.200	
	14	3096	31.441	319.454	8.000	4.484	
	15	3655	31.442	312.986	222.000	220.000	
	16	3458	31.442	311.452	10.000	3.668	
	18	1276	31.405	306.481	-2.000	-0.702	
	19	2643	31.410	305.009	9.000	7.106	
SA-6	1	2127	32.592	299.407	0.000	0.440	
	1A	2127	32.604	299.518	-2.000	-1.978	
	4	1197	32.443	291.496	1.600	0.996	
	5	3021	32.573	291.593	5.000	5.011	
	6	2175	32.565	299.906	12.600	11.100	
SA-7	1A*	5823	32.006	42.987	89.000	88.810	
	2A*	4582	30.024	44.809	1559.000	1554.000	
	3	2676	31.514	141.408	55.000	51.560	
	3A	2702	31.500	141.452	60.000	57.570	
	4	323	31.893	149.323	0.700	0.435	
	4A	338	31.862	149.118	1.700	1.248	
	5	4021	31.710	36.973	-3.000	0.000	
	6	3849	31.752	135.926	-22.000	-17.420	
	7	4573	31.724	135.309	9.000	4.102	
	8	3029	31.721	134.797	7.000	5.021	
	9	1100	31.749	128.814	1.500	1.008	
	10	4238	31.731	128.296	9.000	8.518	
	11	4774	31.724	127.772	-4.000	2.245	
	12	3620	31.730	127.230	10.000	6.663	

TABLE II

Satellite	Pass	No. of Points	i(degrees)	R_{Ω}(degrees)	$10^4 t$	$10^4 f(t)$	Radar
TR-41B	1*	73	32.101	190.577	52.300	52.260	R-113
	1	27	32.421	188.991	0.200	0.007	R-114
MA-5	1	22	32.568	154.441	-0.010	0.000	R-113
	2	20	32.572	153.951	0.050	0.005	
	2	8	32.565	153.847	0.023	0.000	R-114
MA-6	1	18	32.543	233.952	0.006	0.000	R-113
	2	31	32.533	233.446	0.040	-0.001	
	2	4	39.825	262.834	0.047	0.045	R-114
	3	34	32.534	232.901	0.030	0.018	R-113
MA-7	1*	26	33.747	299.009	6.290	6.280	
	2	23	32.547	294.555	0.000	0.011	
	2	7	32.544	294.455	0.002	0.000	R-114
	3	22	32.604	294.557	-0.010	0.000	R-113
SA-5	1	22	31.397	170.797	0.091	0.000	
	3*	21	30.873	151.461	1.960	1.911	
	4	19	31.177	140.479	0.470	0.443	
	5	10	31.105	135.306	0.103	0.101	
	6	17	31.499	133.262	0.030	0.260	
	7	4	32.639	46.736	0.004	0.002	
	8	3	70.877	110.252	0.002	0.002	
	9	6	30.489	22.947	0.029	0.031	
	10	27	31.441	28.655	-0.050	-0.006	
	11	26	31.433	28.435	0.040	0.006	
	13	19	31.746	341.226	0.631	0.625	
	14	26	31.443	319.479	-0.030	0.002	
SA-5	15	31	31.439	312.966	0.030	0.010	R-113
	16	29	31.438	311.479	0.060	0.004	
	18	11	31.405	306.488	0.006	-0.001	
	19	23	31.444	304.827	0.030	0.014	
SA-6	1	18	32.594	299.415	0.000	-0.002	
	1A	18	32.603	299.613	0.000	-0.004	
	4	10	32.613	292.225	0.000	0.000	
	5	26	32.572	291.573	0.010	0.002	
	6	19	32.568	299.894	0.020	0.000	
SA-7	1A*	49	25.889	53.554	1209.000	1209.000	
	2A*	39	30.932	43.219	31.050	31.060	
	3	23	31.433	140.610	0.680	0.685	
	3A	23	31.422	140.684	0.620	0.690	
	5	34	31.708	36.983	0.000	0.000	
	6	33	31.748	135.913	-0.020	0.001	
	7	39	31.724	89.489	0.020	0.024	
	8	26	31.715	134.820	0.024	0.005	
	9	10	31.727	128.748	0.009	0.000	
	10	36	31.733	128.309	0.010	0.001	
	11	40	31.724	127.780	-0.070	0.000	
	12	31	31.730	127.233	0.040	0.001	

REFERENCES

1. *Advanced Calculus*, Taylor, Angus E., Ginn, New York, 1955.

2. *Mathematics for the Physical Sciences*, Wilf, Herbert S., Wiley, New York, 1962.

3. *Numerical Calculus*, Milne, William Edmund, Princeton University Press, Princeton, New Jersey, 1949.

4. *Matrices, Their Meaning and Manipulation*, Bickley, W. G. and Thompson, R.S.H.G., D. VanNostrand Co., Inc., Princeton, New Jersey, 1964.

THE EFFECT OF METRIC CHANGES ON RESOURCE
ALLOCATION DECISIONS

Dr. Richard C. Sorenson
U. S. Army Personnel Research Office

ABSTRACT

Underlying linear programming models is the assumption that the metric of the cost-criterion matrix is an interval scale. The metric of the criterion measures is particularly open to question in the use of the assignment model in manpower resource allocation. The effect on the objective function of using cost/criterion variables of a degenerate metric has been studied by a model-sampling experiment. It is concluded that the assignment model is fairly robust with respect to the metric assumptions generally assumed to be more restrictive.

THE EFFECT OF METRIC CHANGES ON RESOURCE ALLOCATION DECISIONS

Dr. Richard C. Sorenson
U. S. Army Personnel Research Office

One of the central problems in manpower operations research is that of optimal allocation of manpower resources. In this regard a great deal of effort has been expended in an attempt to provide information concerning the degree of achievement or amount of contribution that a person will make in a criterion situation, e.g., job assignment. The use of a predictor variable in estimating the criterion value rather than the criterion variable itself stems from the desire for the information before the person undertakes the proposed assignment. Often a multiple prediction system provides much more reliable estimates of the complex criterion. In recent years considerable success has been attained in the attempt to predict the success of individuals in a particular activity on the basis of a test battery or other objective information. For optimal allocation over a number of criterion activities it is necessary to have not only performance estimates for an individual in a given criterion, but also differential estimates of performance for individuals for a set of criteria, e.g., job categories. However we also find situations where the best estimates of future performance on specific sets of criterion activities are indices, subjective ratings, or rankings (i.e., weakly ordered metrics). Furthermore we find situations where the cost/criterion values are transformed for various management reasons in a non-linear manner. An example of the latter in the manpower field is the non-linear transformation of three-digit Aptitude Area Scores to the one-digit score on which the manpower allocation of enlisted men is based (see AR 611-259).

Of course, the development of differential prediction batteries, rating scales, and performance rankings are solutions to only a part of the optimal allocation problem. Specifically, the problem is to allocate the M available individuals over the N jobs, satisfying certain restraints regarding quotas and minimum qualifications, in such a manner that the objective function (of the cost/criterion values) is maximized. It may be recognized that this allocation problem is a special case of the distribution model in linear programming, and that solutions to it are available (Ford and Fulkerson, 1956; Flood, 1960, 1961; Kuhn, 1955a,b).

It is realized that one of the assumptions under which the linear programming techniques are used is that the cost/criterion matrix is composed of interval-scale variables. Since the metric of the criterion measure is at times open to question, the legitimacy of the use of the assignment model in optimal manpower allocation may be questioned. This is particularly so in complex personnel situations where the performance criterion is not percisely defined. In these situations criterion measures (e.g., test scores, ratings, and rankings) with somewhat arbitrary or artificial scales (i.e., metrics) are used. The question has been raised by King (1965) concerning the use of the

assignment model for the manpower allocation problem where the cost/criterion matrix is composed of performance estimates rather than of measures of performance on an interval-scale metric. On the positive side Brodgen (1955) has shown that as the size of the manpower pool increases, the performance estimates for a set of criterion activities derived by the use of linear multiple regression equations yields, upon assignment of men to jobs, an objective function equal to or higher than that obtained by any other assignment of individuals to jobs that is based on the predictor variables. This is an important conclusion which makes it possible to substitute predictions for actual performance measures, providing an appropriate metric exists. Thus the degree to which the assumption of equal interval scale measures in the cost/criterion matrix of the linear programming assignment model must be met in order to justify the use of the model is a crucial problem.

In order to study the effect of metric changes on the results of optimal allocation an assignment procedure was simulated. At the completion of basic combat training, enlisted men (EM) in the U. S. Army are assigned to advanced individual training, in part, on the basis of results of the Army Classification Battery (ACB). The ACB variables are transformed to Aptitude Area (AA) scores as indicated in Figure 1. Samples from the EM input population were generated as AA score variables with statistical properties of those variables for the actual EM input population. The basic step in the implementation of such a model sampling experiment is the generation of uniformly distributed random numbers. In this experiment computer routines which generate pseudorandom numbers by the power residue method (Larkin, 1965) were used. These distributions of uniform variables are then transformed to distributions of normally distributed variables. This transformation results in a matrix, X, of order n by k, i.e., n entities are represented each by a vector of k simulated scores:

$$X = \begin{bmatrix} X_{11}, & X_{12}, & \ldots, & X_{1k} \\ X_{21}, & X_{22}, & \ldots, & X_{2k} \\ \vdots & & & \\ X_{n1}, & X_{n2}, & \ldots, & X_{nk} \end{bmatrix} \qquad [1]$$

$$\left.\begin{array}{lll} \text{where} & X'X & \to nI \\ \text{and} & 1'X & \to 0 \\ \text{when} & n & \to \infty \end{array}\right\} \qquad [2]$$

ARMY CLASSIFICATION BATTERY

TEST	SYMBOL	TITLE	SYMBOL	FORMULA	OCCUPATIONAL AREAS
VERBAL	VE	INFANTRY - COMBAT	IN	$\frac{AR + 2CI}{3}$	11- Infantry Combat
ARITHMETIC REASONING	AR	ARMOR ARTILLERY	AE	$\frac{GIT + AI}{2}$	12 to 19 -
PATTERN ANALYSIS	PA	ENGINEERS-COMBAT			All Other Combat
CLASSIFICATION INVENTORY	CI	ELECTRONICS	EL	$\frac{MA + 2ELI}{3}$	2-Electronics 3-Electrical Maintenance
MECHANICAL APTITUDE	MA				
ARMY CLERICAL SPEED	ACS	GENERAL MAINTENANCE	GM	$\frac{PA + 2SM}{3}$	4-Precision Maintenance 5-Military Crafts
ARMY RADIO CODE	ARC				
GENERAL INFORMATION	GIT	MOTOR MAINTENANCE	MM	$\frac{MA + 2AI}{3}$	6-Motor Maintenance
SHOP MECHANICS	SM	CLERICAL	CL	$\frac{VE + ACS}{2}$	7-Clerical
AUTOMOTIVE INFORMATION	AI	GENERAL TECHNICAL	GT	$\frac{VE + AR}{2}$	8-Graphics 9-General Technical 0-Special Assignment (part)
ELECTRONIC INFORMATION	ELI	RADIO CODE	RC	$\frac{VE + ARC}{2}$	05-Radio Code

Figure 1. Army Classification Battery (ACB) Tests and Army Aptitude Area Scores as Functions of ACB Variables.

We see then that for each sample we generate a matrix that has as the expectation for its covariance matrix the identity.

Now we desire to further transform the matrix X by post multiplication by a matrix T such that the resulting matrix has for its expected covariance matrix a given matrix C:

$$\left.\begin{array}{ll} & XT = Y \\ \text{where} & Y'Y \rightarrow nC \\ \text{when} & n \rightarrow \infty \end{array}\right\} \quad [3]$$

The matrix C is specified as a function of the desired standard deviation and intercorrelation of the variables:

$$C = s\ R\ s \quad [4]$$

where R is the desired correlation matrix and s is the diagonal matrix of standard deviations.

We wish to find the matrix T such that the conditions in [3] will hold. From these equations we may write the requirement that:

$$\left(\frac{1}{n}\right) Y'Y = \left(\frac{1}{n}\right) T'\ X'XT \rightarrow C \quad [5]$$

when $n \rightarrow \infty$

From [2] we see that

$$\left.\begin{array}{ll} & \frac{1}{n} X'X \rightarrow I \\ \text{when} & n \rightarrow \infty \end{array}\right\} \quad [6]$$

and from [5] and [6] we have

$$T'T = C \quad [7]$$

We may represent the matrix C in terms of its basic structure (Horst, 1965):

$$\left.\begin{array}{ll} & C = QdQ' \\ \text{where} & QQ' = Q'Q = I. \end{array}\right\} \quad [8]$$

We know that the matrix C to any power, e.g., a, may be formed by raising the diagonal matrix, d, to that power, premultiplying by Q and post-multiplying by Q' :

$$C^{a} = Qd^{a}Q' . \qquad [9]$$

thus

$$C^{\frac{1}{2}} = Qd^{\frac{1}{2}}Q' \qquad [10]$$

and

$$C^{\frac{1}{2}}\, C^{\frac{1}{2}} = Qd^{\frac{1}{2}}Q'\, Qd^{\frac{1}{2}}Q' = QdQ' = C. \qquad [11]$$

We will let

$$T = C^{\frac{1}{2}} \qquad [12]$$

so that equation [7] is satisfied:

$$T'T = C^{\frac{1}{2}}\, C^{\frac{1}{2}} = C. \qquad [13]$$

Hence a transformation solved for by equation [10] meets the requirement of [7] and while there are an infinite number of transformations that meet this requirement, the one indicated is by far the most advantageous since it provides for uniformity of rounding errors and impartially improves normality of the transformed scores. Thus, samples of personnel may be simulated by building into the score distributions the statistical characteristics of the variables of interest -- in this case the Aptitude Area score variables. Tables 1 - 3 indicate the matrices which apply to the variables in question. An additive constant was used in the generation of the scores so that the expectation for the mean of each variable would be 100.

In this manner we could sample from a population, form the cost/criterion matrix which for the purposes of this experiment we assume to be composed of variables of an appropriate metric, and perform the experiment outlined below. The purpose as indicated was to study the effect of change of metric on the results of the optimal allocation procedure. As the samples were generated the variables were further transformed successively by each of eight transformations. The optimal allocation was accomplished by a computerized algorithm based on the Hungarian Method (Kuhn, 1955a). The quota restrictions imposed were uniform over the eight job categories, as preliminary research has indicated that perturbation of the quotas within range of Army usage does not affect the relationships among other variables. The objective function was then calculated using the assignment matrix and the original cost/criterion values which had means of 100 and standard deviations equal to 20. In this manner the amount of loss due to inferior metrics of the other criterion measures will be comparable over the eight transformations. The values reported are average values over

Table 1

Covariance Matrix, C, for Aptitude Area Score Variables

1.00	0.65	0.63	0.67	0.57	0.66	0.76	0.68
0.65	1.00	0.76	0.77	0.90	0.57	0.64	0.60
0.63	0.76	1.00	0.78	0.79	0.58	0.68	0.61
0.67	0.77	0.78	1.00	0.74	0.66	0.73	0.65
0.57	0.90	0.79	0.74	1.00	0.48	0.55	0.51
0.66	0.57	0.58	0.66	0.48	1.00	0.87	0.83
0.76	0.64	0.68	0.73	0.55	0.87	1.00	0.87
0.68	0.60	0.61	0.65	0.51	0.83	0.87	1.00

Table 2

Covariance Matrix, C, for Aptitude Area Score Variables

309.4081	201.9156	202.5735	212.6068	183.2808	199.6817	247.9838	209.9191
201.9156	311.8756	245.3468	245.3115	290.5423	173.1386	209.6595	185.9598
202.5735	245.3468	334.1584	257.2215	263.9851	182.3613	230.5839	195.6965
212.6068	245.3115	257.2215	325.4416	244.0307	204.7901	244.2887	205.7913
183.2808	290.5423	263.9851	244.0307	334.1584	150.9197	186.5017	163.6151
199.6817	173.1386	182.3613	204.7901	150.9197	295.8400	277.5822	250.5438
247.9838	209.6595	230.5839	244.2887	186.5017	277.5822	344.1025	283.2307
209.9191	185.9598	195.6965	205.7913	163.6151	250.5438	283.2307	308.0025

Table 3

Matrix $C^{1/2}$ used in Transforming Variables to Represent Aptitude Area Scores for an EM Input Population

14.1889	3.8009	3.5401	3.8875	2.9991	3.6689	5.2631	3.9786
3.8009	12.7125	4.7210	4.9369	7.7572	2.7397	3.4140	3.1380
3.5401	4.7210	14.0658	5.3884	5.9798	2.8229	4.2470	3.2697
3.8875	4.9369	5.3884	13.6033	4.9263	3.7826	4.6312	3.4428
2.9991	7.7572	5.9798	4.9263	13.7632	1.9635	2.5158	2.3095
3.6689	2.7397	2.8229	3.7826	1.9635	12.9704	6.7255	5.9372
5.2631	3.4140	4.2470	4.6312	2.5158	6.7255	12.9796	6.7262
3.9786	3.1380	3.2697	3.4428	2.3095	5.9372	6.7262	13.1899

the several samples generated.

The eight transformations that were studied were:

A. Additive transformation to two digit score with possible values of 0 - 99. This was done by subtracting 50 and truncating. Remaining values less than 0 and/or greater than 99, if any, were replaced by 0 and 99 respectively. This transformation would provide the most likely candidate for the formation of a two digit variable. It was felt that the non-linearity in the transformation at both tails would be negligable; however, some information is lost in the dropping of all but 2 significant digits and in the truncation at the tails of the distribution.

B. Gross linear transformation to one digit score. This transformation was accomplished by subtracting 50, dividing by 10 and dropping all but one digit. Again, those values less than 0 and greater than 9 were replaced by 0 and 9 respectively. It was thought that it would be of interest to compare this one-digit transformation with non-linear one-digit transformations including that which is currently in use in the U. S. Army.

C. Linear transformation for the top half of the distribution. In optimal allocation it is sometimes assumed that the preservation of information in the top half of the score distribution is more important than in the lower half of the distribution. Hence this transformation was formed by subtracting 95, dividing by 5 and dropping all but 1 digit. All values less than 100 were transformed to 0. Entries transformed to values greater than 9 were replaced by 9.

D. Non-linear transformation with smaller intervals in the center of the distribution. Consideration of information measures indicate that more information is preserved when the resulting values appear with approximately equal frequency. When dealing with normally distributed variables, this consideration dictates the use of small intervals for the center of the distribution and larger intervals for the tails. The following transformation was used:

-	69.9̄	=	0
70 -	79.9̄	=	1
80 -	89.9̄	=	2
90 -	94.9̄	=	3
95 -	99.9̄	=	4
100 -	104.9̄	=	5
105 -	109.9̄	=	6
110 -	119.9̄	=	7
120 -	129.9̄	=	8
130 -		=	9

E. Non-linear transformation with smaller intervals for the upper half of the distribution. As indicated in C, preservation of information in the top half of the distribution may be more important than elsewhere in the distribution. The following transformation was used:

-	79.9̄	=	0
80 -	99.9̄	=	1
100 -	109.9̄	=	2
110 -	114.9̄	=	3
115 -	119.9̄	=	4
120 -	124.9̄	=	5
125 -	129.9̄	=	6
130 -	134.9̄	=	7
135 -	139.9	=	8
140 -		=	9

F. Ranking of jobs within individuals. The best available information is sometimes of the ordinal type, e.g., ranking of the jobs for each individual. This is strictly a weakly ordered metric.

G. Ranking of individuals within jobs. In this transformation the lowest estimate of performance on each job was replaced by 1, the next lowest by 2 ...., and the highest performance estimate for each job was replaced by N.

H. Current score transformation (AR 611-259):

-	$79.\bar{9}$	=	0
80 -	$89.\bar{9}$	=	1
90 -	$94.\bar{9}$	=	2
95 -	$99.\bar{9}$	=	3
100 -	$104.\bar{9}$	=	4
105 -	$109.\bar{9}$	=	5
110 -	$119.\bar{9}$	=	6
120 -	$124.\bar{9}$	=	7
125 -	$129.\bar{9}$	=	8
130 -		=	9

The results are indicated in Table 4 in terms of the allocation average, i.e., the average original cost/criterion value for the individuals on the job to which they are assigned. We note that 4000 entities were simulated and grouped into subsamples in three different ways: I) 250 samples of size 16, II) 100 samples of size 40, III) 20 samples of size 200. The base metric with which we will make our comparisons is transformation A. This transformation is analogeous to that involved in rounding the criterion values to three digits (where the distribution is truncated to provide a range of 50 - 150 this is equivalent to a scale from 0 - 99). In making our comparisons we should recall that the allocation average under random assignment would be 100. Thus by optimally allocating using metric A the gain over random assignment is 13.76 units for samples of size 200. Using the weakly ordered metric, F, the gain over random assignment was 13.39 units, a 3% loss from the base metric. Relatively little reduction in the allocation average results from the use of the ordinal metric. Also from Table 4 we see that metric C was most inferior. In this case the gain over random assignment is 18% less than for the base metric.

Our purpose was to estimate the magnitude of the differences between the results for the different metrics in terms of the average criterion values for the jobs to which the simulated individuals were assigned. Some methods were known to be superior to certain others, and, since we are not sampling from a universe in which the alternative methods could conceivably be homologous, a difference which is not large enough to be statistically significant can not be presumed to be zero. Rather, our best guess as to the magnitude of that difference must necessarily be the obtained, albeit non-significant, results. Thus the testing of significance of differences was secondary to estimating direction and magnitude of differences. For this reason the same sample of simulated

Table 4

Results of Three Replications of the Model Sampling Experiments 1/

Metric 3/		Replication 2/		
Identification	Discription	I	II	III
A	Base Metric	112.68	113.36	113.76
B	1 digit, symmetrical	111.89	112.48	112.68
C	1 digit, upper half	110.50	111.04	111.30
D	1 digit, center intervals smaller	111.93	112.48	112.78
E	1 digit, upper intervals smaller	111.37	112.07	112.26
F	ranking within individuals	112.02	112.79	113.39
G	ranking within jobs	111.32	112.59	113.50
H	current operational method	111.87	112.47	112.73

1/ The experiment was replicated three times using the same sample but with the optimization over different sizes of subsamples.

2/ I: 250 subsamples of size 16, II: 100 subsamples of size 40, III: 20 subsamples of size 200.

3/ The 8 transformations, A - H, are described in the text.

individuals was assigned under all conditions (i.e., metrics and sample sizes), a procedure which reduces the chance factor in our measures of the difference between conditions at the cost of not being able to say explicitly how much the chance variation has been reduced. While we are not willing to admit the possibility that the differences between the effect of the various conditions could be zero in the universe, we must consider the possibility that the ordering indicated by our samples could be due to chance. Thus, it is of interest to determine whether the ordering indicated in Figure 2 could be the result of sampling fluctuations. Tests of the significance in difference in performance means using the appropriate tests for correlated variables, resulted in the rejection of the null hypothesis at the .05 level for those cases not bracketed. Note then for instance that the difference in results using metrics E and G were not significant at the .05 level for samples of size 16 when only 250 samples were used. Of course as in any statistical testing this evidence does not establish the hypothesis of no difference, but merely indicates that confidence (at the indicated level) may not be placed in the ordering. This we have indicated in Figure 2 by enclosing D, B, and H. In each case these three are significantly different from the remaining metrics but not different from each other (except for case I where D was significantly different from B). It is realized that one must be cautious in interpreting multiple significance tests, however the technique is satisfactory for the present use. It is of interest to note the change in the relative position of metric G. As would be expected, as the sample size increases, the ranking of individuals within jobs provides an increasingly satisfactory metric. We may note that for medium or large samples (200 or larger) the ranking methods provide metrics for criterion values in the assignment model which are nearly as efficient as the base metric. It is felt that the ranking of jobs within individuals would be even more satisfactory as the number of jobs increases. This method should be considered as a possible way of obtaining criterion values in specific situations if some degradation of the metric is necessary for whatever reason.1/ It was shown that the methods E, C which preserve greater information in the upper portions of the criterion distributions at the expense of information in the lower half of the distribution were clearly inferior.

A word should be said concerning the results for the different sample size. This was of interest primarily in terms of metric G. Other comparisons are available, however. Table 4 indicates that the amount lost by rounding to one digit from two digits with samples of size 200 is approximately equivalent to that lost by reducing the size of samples used for allocation from 200 to 16 while retaining 2 digits of information.

To summarize we may say that results indicate that for some important instances of deviation from the interval scale in the cost/criterion matrix in optimal manpower resource allocation little is lost through the use of a degraded metric. Specifically, the results 1) underline

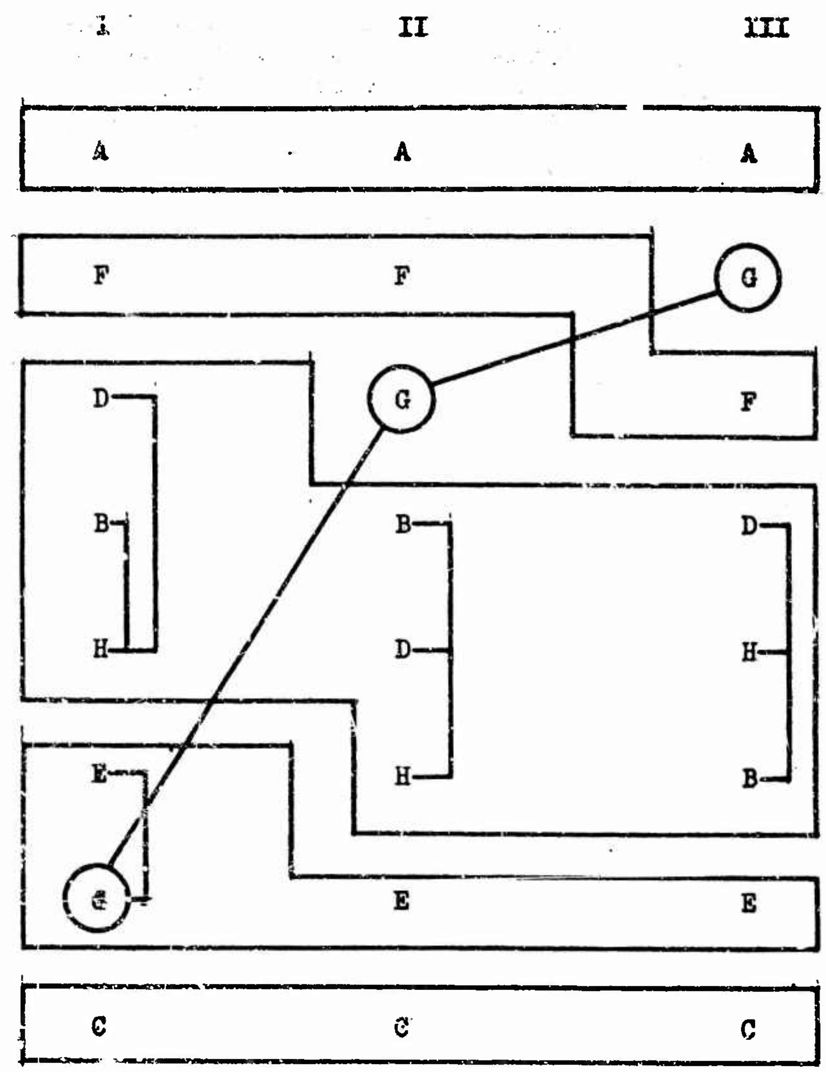


Figure 2. Ranking of the transformations in terms of the results of the model sampling experiment. The enclosed transformations are significantly different from each other, e.g., results with transformation A are significantly different from those with F for all three replications. The bracketed transformations were not significantly different from each other at the .05 level.

the necessity of preserving information throughout the entire range of the distribution 2) provide an estimate of the loss due to allocating on the basis of small manpower pools, and 3) demonstrate that the personnel assignment model is quite robust with respect to the metric assumptions generally assumed to place severe restrictions on the use and interpretation of model results.

1/ It may be less expensive to determine the ranking of jobs within individuals in terms of the criterion without measuring the criterion outright.

REFERENCES

Brogden, Hubert E. Least squares estimates and optimal classification, Psychometrika, 20:3, 1955.

Flood, Merrill M. An alternative proof of a theorem of Konig as an algorithm for the Hitchcock distribution problem, in Richard Bellman and Marshall Hall, Jr. eds., Proceedings of the Tenth Symposium in Applied Mathematics, 1960.

Flood, Merrill M. A transportation algorithm and code, Naval Research Logistics Quarterly, 8:3, 1961.

Ford, L. R. and Fulkerson, D. R. Notes on Linear Programming -- Part XXXIII: Solving the transportation problem (RM-1736), The Rand Corporation, 1956.

Horst, Paul. Factor Analysis of Data Matrices, Holt, Rinehart and Winston, Inc., 1965.

King, William R. A stochastic personnel-assignment model, Operations Research, 1965.

Kuhn, H. W. The Hungarian method for the assignment problem, Naval Research Logistic Quarterly, 2:1, 1955a.

Kuhn, H. W. A combinatorial algorithm for the assignment problem, mimeograph, 1955b.

Larkin, W. L. Computer generation of random numbers, Research Memorandum 65-3, U. S. Army Personnel Research Office, 1965.

AN EXPERIMENTAL FIXED INCREMENT WAR GAME

by

Drs. F. J. Murray, W. E. Sewell, R. E. Chandler, and L. E. Winslow
Special Research in Numerical Analysis, Duke University
Cognizant Agency: Army Research Office - Durham

ABSTRACT

This paper describes the structure of a computerized Experimental Fixed Increment War Game designed to explore and test ideas for such games, with emphasis on a military analysis appropriate for programming. In Part I we develop a basic game with no consideration of terrain, and in Part II we discuss the introduction of terrain into the basic game.

AN EXPERIMENTAL FIXED INCREMENT WAR GAME

by

Drs. F. J. Murray, W. E. Sewell, R. E. Chandler, and L. E. Winslow
Special Research in Numerical Analysis, Duke University
Cognizant Agency: Army Research Office - Durham

PART I: General Structure of the Game

1. Introduction. War gaming, of course, is an old vocation and avocation and has gone through many stages down through the centuries. The latest stage brings in the computer and uses it to varying degrees in assisting with the details of the game. The present complexity of the organization and equipment of armed forces certainly makes the computer a welcome addition to the control room of a war game. The computer can do much of the arithmetic, of which there is plenty to be done, and furthermore doesn't make many mistakes. A game in which the computer is used to add and subtract and to keep the books and records, so to speak, is called a computer-assisted game.

We are interested here in what we will call a computerized war game--that is, a game where certain military units are actually played by an electronic computer. Our interest in the subject stems from our long association with STAG, the U. S. Army Strategy and Tactics Analysis Group, and we are indebted to the entire staff of STAG, and particularly to Lt. Col. Charles R. Roberts, Maj. Charles G. Roebuck, Mr. John A. Albertini, Mr. Paul Dunn, and Mr. David R. Howes. Drs. Murray and Sewell have been working with STAG about four years, Dr. Chandler more than two years, and Dr. Winslow about a year. STAG's mission is to develop computerized war games at various levels which will serve to test strategic and tactical plans. STAG has made remarkable progress in training the computer to command military units but like most commanders the computer is not perfect and STAG is always looking for ways to improve its computerized games.

Actually, this particular game was developed for the express purpose of testing ideas, many of which arose during our discussions at STAG. The necessity to be continually operational makes it difficult for STAG to experiment with new ideas, whereas we have no operational responsibility. In fact, we would like to make it clear that this is not an operational game suitable for military application of any type. We are primarily interested in ground combat but hope to keep the structure broad enough to introduce pertinent air and naval action and to test ideas and procedures related thereto. Also, we expect to consider facets of all three types of war games, operational, training, and research [1, p. 1]*. An excellent survey of existing war gaming in our military establishment may be found in

*The numbers in brackets refer to the bibliography at the end of the paper.

[2], and a description of a STAG game was presented to an earlier session of this Symposium [3].

Let me say a word here about some advantages of a computerized war game. Of course, the computations are based on rules and these rules must be decided upon beforehand and they are rigid; i.e., they cannot be changed during the game. This makes the computerized game more difficult to set up but it does provide a means of faithful duplication. Furthermore, the computer makes very few mistakes and can be programmed to keep whatever records may seem desirable. And, in our opinion, these two features are the great advantages of the computerized game--excellent records are available for a detailed analysis of the play, and faithful duplication is attainable. The latter means, for example, that you can change the rate of fire of one type of weapon and play the game with just that one parameter changed. The outcome will then be a measure of the effect of just that change. As we all know, it is almost impossible to duplicate a manual game--you change an umpire or a controller and you change the game. Judgment is inherently a part of a manual game, and the judgment of individuals differs. Frequently the purpose of a game is to determine the effects of changes rather than absolute values, and in the computerized game you do have absolute control over the changes.

Another advantage of the computerized game is that a great deal of computation can be accepted and you know that it will be done correctly. For example, you don't have to aggregate fires and fire effects in order to avoid computation; you can compute the total rounds of each type and handle rather long arithmetic calculations relatively easily and rapidly. Furthermore, calculations can be made at short intervals--the computer doesn't get bored and loaf on the job.

A disadvantage of the computerized war game is that all the problems have to be anticipated, and solved, and there is no room for exception. You can, of course, provide for different conditions beforehand but you cannot count on the computer using good judgment in a situation which was not anticipated. Consequently, you can never hope to have the flexibility (this may be a poor choice of words) in a computerized game that you have in a manual game. On the other hand, this so-called flexibility may not be a real advantage in that it introduces subjective considerations that are often questionable, and certainly not universally accepted, and tends to make war more of an art than a science.

In a computerized war game you must rule out the art, at least on the level of the military units and related data in the computer, and accept the assumptions, the rules, and the formulas prescribed. Therefore it is essential that the military analyst and the players know these assumptions, rules, and formulas, and that they be acceptable to them; in other words, that they yield results which are reasonable and compatible with average military experience.

Since our game is aimed at testing ideas which will contribute to a practical working game, we have concentrated on the military analysis and on "reasonable" assumptions and results and have aimed at the simplest

structure which promises to provide a comprehensive test of the ideas under consideration. The emphasis in this discussion is on a military analysis which can be programmed, but not on the program itself.

2. Game Structure. Our game consists of various subsections which are independent of one another in the sense that each subsection has as its final result a number. While this number might be used by subsequent subsections, altering the method by which this number is obtained does not necessitate altering any other subsection.

The game simulates a conflict between two military forces. These forces consist of a certain number of combat units (basic elements) with specific characteristics, which are subject to orders by "players" who take the part of commanders and staffs at higher echelons. The structure of the game is determined by two factors:

a) the characteristics which describe a basic element;

b) the procedures for modifying these characteristics due to interaction of the elements, or to player action on the elements.

We have taken the element, a military unit, in the computer to be roughly a brigade. The characteristics which specify an element are the quantities of various weapons and vehicles, the location of the element on the map (or ground), the STATE (roughly the mission--this will be more precisely defined later), speed and direction of movement, fire capability in rounds per minute of various classes of weapons, and the total fire being received. Each of these characteristics is computed in an individual subsection of the game.

The characteristics of an element are divided into two classes--controlled and uncontrolled. Once the game has started, the uncontrolled variables are determined by the rules of the game; while the controlled variables are subject to manipulation by the players. The players in this game can issue orders concerning missions, and the rules of the game determine the effects of these orders.

In the modification of the characteristics we assume that a time interval or increment can be chosen so small that the changes in the characteristics of the element can be considered to be instantaneous and occurring at the end points of the interval, and hence, these characteristics are considered to be constant within each time interval. This is our definition of a "fixed increment war game". Clearly, the increment must be small in order for this assumption to be tenable, because events happen fast on the battlefield. But the computer is capable of doing a lot of arithmetic in a short time, much more than could possibly be anticipated in a manual game. For this reason, among others, the manual game is usually a "critical event game". An important consideration in any game is the relation between game time and real time and the computer enables us to play a fixed increment game with a small interval and still keep the game and real times within an acceptable ratio.

The formulas for calculating the changes in the characteristics are the simplest arithmetic expressions which provide reasonable results, from a military point of view, for the specific situation under consideration. The attrition formulas are essentially simplified forms of the Lanchester equations [4]. We recognize that these formulas are only approximations and are limited to a restricted spectrum of situations. This is the price that we pay to keep the game simple. More sophisticated formulas could be incorporated rather easily and without changing the structure of the game.

3. Outline of the Game. A general "Influence Diagram" for the game is shown in Figure 1 on page 5.

To determine the characteristics of the elements at time t we assume that the characteristics of the elements are known at time $t - \Delta t$, where Δt is the time increment.

A brief description of the calculations involved in the Influence Diagram is perhaps in order. Calculation [1] determines the new location from the old location and the speed and direction of movement. Calculation [2] determines the distances between all elements, the distance between each unit and its nearest enemy unit (or units if several are at the same minimum distance), the number of enemy units within small arms range, and a constant used to apportion fire among the various enemy units within range of any weapons class. Calculation [3] determines the attrition for each class of weapons and vehicles; this attrition is dependent upon the time increment Δt, the total enemy fire received from each weapons class, speed of the unit and its STATE. (We assume that a unit in a prepared defensive position, for example, suffers fewer losses than an attacking unit, and a unit moving in convoy at a high rate of speed suffers more losses than one which is dispersed and advancing slowly and cautiously.)

Calculation [4] determines the new STATE of the element. In general, STATE denotes the activity of an element compatible with its mission. There are seven STATES: STATE 1: administrative movement (no enemy threat); STATE 2: contact (enemy under observation but out of firing range); STATE 3: attack; STATE 4: defend; STATE 5: support; STATE 6: break (element no longer combat effective due to excessive losses); STATE 7: withdraw. The concept of STATE is a convenient tool for recognizing the obvious fact that the firepower, attrition, etc., of an element is definitely dependent upon its military activity (STATE). The calculation is based on the old STATE, strength, mobility, and location of the element and the location of the opposing elements.

Calculation [4*] can modify [4] on the basis of orders issued by the players. Such orders specify an objective and a STATE for the unit.

Calculation [5] determines the speed and direction of the element on the basis of the STATE, mobility, location of opposing elements, and the objective. Calculation [6] determines the firepower potential of the element in terms of rounds per minute for the individual weapons classes based on the speed of the element and the fire it has received during the preceding

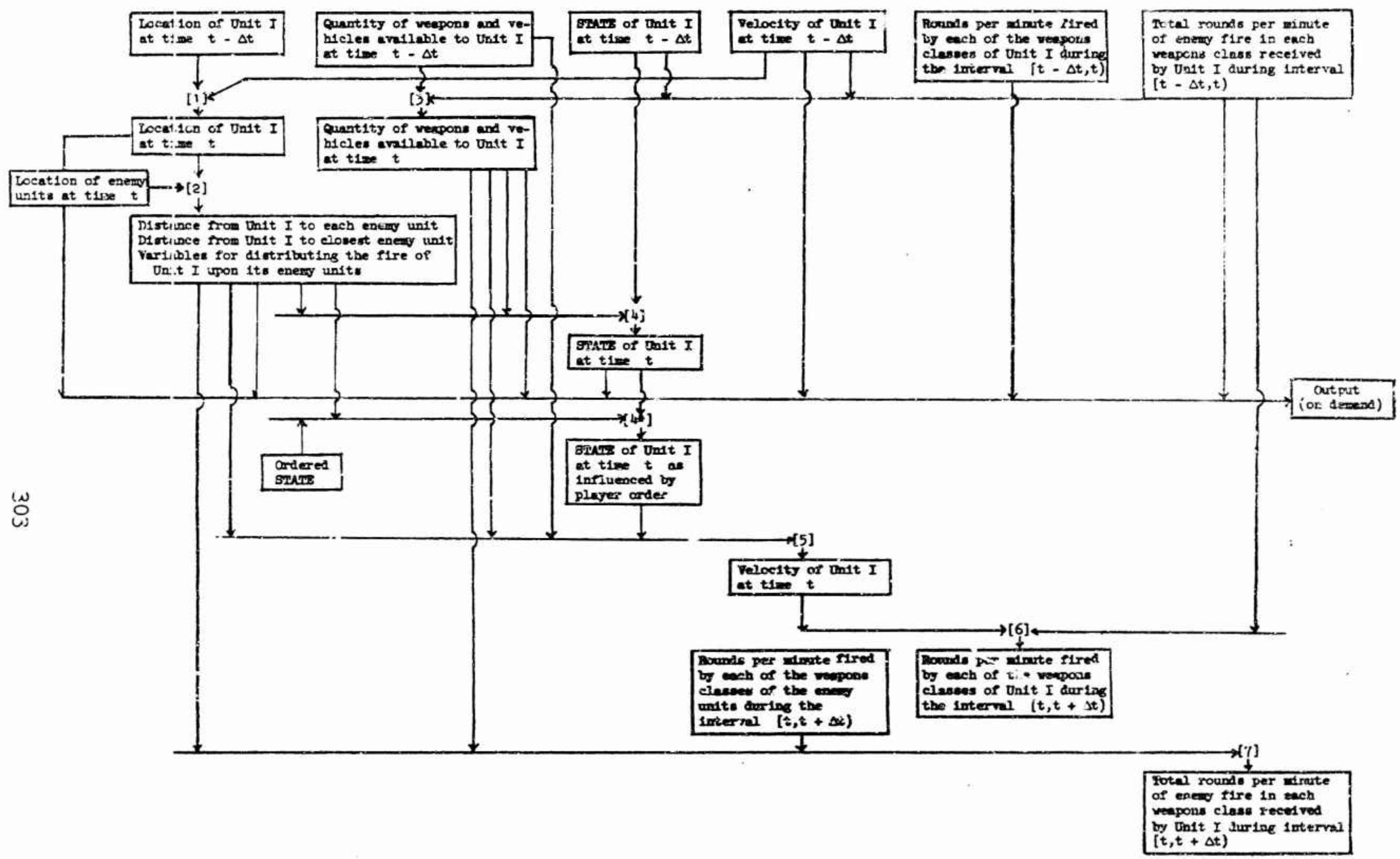


Figure 1. INFLUENCE DIAGRAM

time interval. Calculation [7] determines the total rounds of fire received from the various weapons classes of opposing units (or elements).

The formulas are rather numerous and long and no attempt will be made here to do more than look at two samples; the first serves to indicate the notation, and the second is characteristic of the majority of the formulas used in the game.

Calculation [1] is simple and gives us the new location (location at time t of unit i):

$$X_{i,t} = X_{i,t-\Delta t} + (\Delta t)(S_{i,t-\Delta t})(\cos \theta_{i,t-\Delta t}),$$

$$Y_{i,t} = Y_{i,t-\Delta t} + (\Delta t)(S_{i,t-\Delta t})(\sin \theta_{i,t-\Delta t}),$$

where $(X_{i,t-\Delta t}, Y_{i,t-\Delta t})$ denotes the coordinates of unit i and time $t - \Delta t$; $\theta_{i,t-\Delta t}$ the azimuth on which unit i is moving during the interval $[t - \Delta t, t)$; $S_{i,t-\Delta t}$ the speed in kilometers per minute at which unit i is moving during the interval $[t - \Delta t, t)$; and Δt the length of the time interval $[t - \Delta t, t)$ in minutes.

Calculation [3] gives the loss in weapons and vehicles suffered by unit i during the interval $t - \Delta t, t$. A sample formula is:

$$W1_{i,t} = W1_{i,t-\Delta t}[1 - \{(EKC11)(\Delta t)(TFR1_{i,t-\Delta t}) + (EKC12)(\Delta t)(TFR2_{i,t-\Delta t})\}SLR1]$$

$$\text{where } SLR1 = \begin{cases} 0, & STATE_{i,t-\Delta t} = 1 \text{ or } 2 \\ 1, & STATE_{i,t-\Delta t} = 4 \text{ or } 5 \\ \dfrac{STC11}{STC12 + S_{i,t-\Delta t}}, & STATE_{i,t-\Delta t} = 3, 6 \text{ or } 7 . \end{cases}$$

Here $WK_{i,t-\Delta t}$, $K = 1$ in this case, denotes the number of weapons of class K available to the i-th unit during the interval $[t - \Delta t, t)$; specifically, we treat W1 as small arms, W2 as artillery, and W3 as tanks. EKCIJ denotes the effective kill coefficient on one weapon of class I fire by one round of class J fire, $I, J = 1, 2, 3$. $TFRK_{i,t-\Delta t}$ denotes the total rounds per minute of class K fire received by the i-th unit during the time interval $[t - \Delta t, t)$, $K = 1, 2, 3$. STC11 and STC12 denote constants depending on the STATE used in the formulas for SLRK, K = 1, 2, 3, V. SLR1 is a constant defined by the formula above. This indicates the wide use we make of constants and the freedom with which we alter the constants to fit varying situations.

In connection with Calculation [5] we gave the computer a choice. If $STATE_{i,t} = 3$ we use the losses in manpower (W1) over the interval $[t - \Delta t, t)$ as a loss rate and calculate the expected losses to be suffered during the time required for the unit to move over a distance equivalent to its distance from the nearest enemy unit. The expected losses would then be given by

$$EL_{i,t} = \frac{DMIN_{i,t}}{(S_{i,t-\Delta t})(\Delta t)} [W1_{i,t-\Delta t} - W1_{i,t}] ,$$

where $DMIN_{i,t}$ denotes the minimum distance from the i-th unit to the nearest enemy unit. If $W1_{i,t} - EL_{i,t}$ is less than 400 (the break level for a unit) then the unit is receiving too much fire to proceed further and so $S_{i,t} = 0$. If $W1_{i,t} - EL_{i,t}$ is greater than or equal to 400, the unit can proceed further and so $S_{i,t} \neq 0$.

Also, we adjust the speed to prevent one unit from colliding with an enemy unit. This is done by requiring the distance traversed from time $t - \Delta t$ to time t to be less than one-half the distance from unit i to the nearest enemy unit, if an enemy unit is in sight.

4. Program of the Game. This game has been programmed in BASIC FORTRAN. It used about 5000-words of storage on the IBM 7072 for a total of fifty elements (units). Its running time is output-bound on this particular machine, indicating that it could be run on a much smaller machine providing the machine had sufficient storage.

For the game we used a map 100 kilometers square, and limited the number of units to 25 blue units and 25 red units. We gave each unit three classes of weapons, 1000 rifles, 25 artillery pieces, 25 tanks, and 500 vehicles. These numbers are arbitrary; however, the formulas are dependent to some extent upon the numbers. We found, for example, that if we gave a unit 10,000 rifles the formulas did not work out very satisfactorily. We can use the same basic formulas but the constants have to be adjusted.

With regard to capabilities, we assumed that a unit could see anything within 15 kilometers and could fire on a target at 10 kilometers or less. We took the rifle range to be 300 meters. Rates of fire and movement were scaled down from basic rates by the situation; i.e., STATE, number of vehicles available, etc.

As far as firing was concerned, a unit automatically fired on a unit within range and allocated its fire between targets according to range.

We can require output at regular intervals, each time interval, every ten time intervals, etc. For input, we have to stop the machine. Our output consisted of the identification of the unit, its location, STATE, speed, objective, locations of enemy units within sight, weapons and vehicle

status, fire being delivered in rounds per minute, and the fire being received in rounds per minute.

5. Some Basic Features of the Game. In the development of this game we have relied on the simplest mathematical relations which will reasonably depict ~~the results rather~~ than attempting to develop a complete mathematical model which would be amenable to optimization techniques. At the same time, we have insisted on mathematical relationships sufficiently precise and comprehensive to serve as a framework for testing ideas that can or could be useful in a working war game. The Influence Diagram shows that the program is built of subsections, the interior workings of which are mutually independent. Thus, even though we have employed very simple mathematical relations, inside the individual subsections it is easy to replace them with more sophisticated formulas without changing the general structure of the program.

The length of the time increment is a vital factor in any game of this type based on difference equations. Consequently, we have studied the variation of results caused by changing the length of interval. We have found, for example, that for our particular formulas a time interval of one minute produces the most realistic results. For an interval much smaller than one minute round-off begins to introduce appreciable errors into the result. Since we are assuming that the values of the characteristics of the element are constant in the time interval, an interval much larger than one minute invalidates this assumption for actual combat.

To make these simulations work we needed a military unit which was more or less homogeneous and which operated tactically as an entity. Thus, we have selected a unit which roughly corresponds to an infantry brigade. The unit (or element) can be altered in composition within limits, we have found, and the same formulas apply; however, the composition cannot be changed radically without changing the formulas.

In calculating firepower and attrition by firepower we have used rounds per minute of the separate weapons classes as units of measure, rather than attempting to aggregate several fires (small arms, artillery, etc.) into a single number. The desirability, in fact, the necessity, for aggregating in a manual war game is obvious but has never been satisfactory. With the computer it is not necessary to aggregate and by not aggregating we recognize the well-known fact that round-per-minute of various types of fire produce different and distinct effects on the men and weapons of the military unit.

6. Problems to be Investigated. The game as described above provides the basic framework for a study of specific concepts relative to their value in a practical, usable war game. We have, in fact, already studied to some extent various ideas and found the structure adequate for our purposes. We propose to investigate the most effective methods of introducing additional factors with emphasis on those which are essential to the development of a practical war game.

The basic game assumes a flat terrain with constant mobility

index, target detection, firing range, and vulnerability, and our game develops from the action of opposing units moving toward each other and thereby producing a meeting engagement. The importance of terrain variation in ground combat is universally recognized; in fact, it is just as much a part of a military unit's combat potential as its men, its weapons, or its vehicles. Consequently, we expect to devote our first effort to its introduction, and we anticipate a continuing emphasis on refining our treatment of terrain.

Military intelligence is also an important part of a unit's combat potential, and in the course of our investigations we hope to introduce this factor, especially for larger units, as a characteristic of the unit affected by information from the flank (adjacent units) and from the rear as well as by observation to the front.

A related problem is the coordination of friendly units. The present game handles each unit more or less as a separate entity and does not tie adjacent units together as a team. We hope to introduce not only the interchange of information but also the coordination of fire and movement between adjacent units.

We recognize the fact that our equations are rough approximations, and we have already found that the equations are definitely limited in scope relative to sizes of units and situations. We will, of course, endeavor to refine these equations and to modify the related parameters in such a way as to produce more realistic results. However, we do not consider ultimate refinement of specific formulas highly important or particularly useful from a practical point of view. We believe that the main value of a computerized war game consists of the ability to detect relative variations and to study the effect of varying one or more parameters with the assurance that we are actually observing the effect of varying just those parameters. In a manual game individual judgments and decisions which inevitably enter do not permit this exact repetition of results.

In our game we assume that a unit (element) occupies a circular area of radius 1/2 kilometer. Actually, most military units in a combat zone are deployed over an area whose shape and orientation depend upon the situation. We propose to study the introduction of this shape and orientation and its effect on the fire and combat power of the unit.

Another simplifying assumption in our present game is that no battle lasts long enough to generate any significant logistical problems. With the real time to game time ratio which we have at present, we see no difficulty in extending game time to a period which will involve logistics as a major factor.

This game was originally conceived as a vehicle for investigating the potential war games which can be played on computers. We intend to study the memory requirements and various techniques of reducing these requirements, the ratio of game time to real time, communications between players and the computer, etc. Another important facet of this work is the study of various compiler languages and their applicability to war gaming.

Most of the work in this area has been done in one or another dialect of FORTRAN; the limitations of FORTRAN are obvious to anyone who has attempted the computerization of a war game.

As mentioned previously, one of the most interesting problems is the development of an effective game for a CPX, on the Division level, for example. This could probably be imposed on the present game without too much difficulty and programmed in FORTRAN, thus making it available to many Army units, especially National Guard and Reserve. For such units the computer would replace umpires and controllers and thus make the CPX a much more effective training vehicle.

PART II: Introduction of Terrain

1. General Principles. We are concerned here with the introduction of terrain into the basic game and the study of its effect on combat.

From a tactical point of view the terrain features which need to be considered depend upon the size of the unit. For example, a small patch of woods may be vitally important to a platoon but can be ignored as far as a brigade is concerned. We use a military map of the area, containing just those terrain features of tactical significance to the type of unit (a brigade) under consideration. In our treatment we will endeavor to make the procedure as simple as possible, using regular geometric figures for boundaries, and indices for such factors as mobility, visibility, firepower, and vulnerability.

2. Methods of Handling Terrain. Various methods of treating terrain in computerized war games have been employed. The method of indexing by grid squares has the advantage of being easy to implement and requiring very little computation. On the other hand, it uses a large amount of storage space. Thus, this method is only to be recommended for those war gamers who have a large amount of computer storage space available.

A modification of the above method is indexing squares by exception; i.e., all squares except those associated with specific terrain features are considered homogeneous. If only a few squares are exceptional, this method is both fast and uses very little storage. But in the general case this method uses less storage space and more time, and a careful analysis is essential to determine the preferred method for a specific game and machine.

Continuing the concept of trading space for time, we may index by areas of exception. The critical consideration here is the way in which the areas are described. There are various methods of describing these areas and the simplest methods involve the use of regular geometric figures to describe the boundaries; e.g., the rectangle, the circle, the ellipse, the lens, etc. Each class of geometric figures gives rise in turn to a particular procedure with its own advantages and disadvantages.

A problem associated with indexing by areas of exception is comparing the location of each unit with the location of each terrain feature. This consumes considerable time, but some time can be saved at a small cost in storage by dividing the area into rectangular sub-areas so that each sub-area intersects a small fraction (say, one-fifth) of the total number of terrain features. Then it is only necessary to compare the location of each unit with the location of those terrain features which intersect the sub-area containing the given unit. The appropriate choice of the number of sub-areas depends upon the number and size of the terrain features.

3. Numerical Quantities Associated with a Terrain Feature. Once a method has been determined for describing the location and shape of a terrain feature, one must analyze from a computational point of view the effect of the terrain feature on the tactics of the unit. It is to be noted that we consider only those terrain features in the area which affect the tactics of the basic element (brigade).

A terrain feature, in general, affects the tactics of a unit as soon as it can see the feature. At this time the unit moves with more caution and increases its patrol activity, since the feature may contain an enemy. Consequently, we find it convenient to introduce the concept of a tactical perimeter roughly 15 kilometers (level range of visibility) from the boundary of the feature, and the concept of a tactical fringe consisting of the area between the tactical perimeter and the boundary of the feature.

Thus we associate with each terrain feature a Speed Degrading Factor-Terrain (SDFT); as soon as the unit comes within sight of a terrain feature (crosses the tactical perimeter) we degrade the speed of the unit by this factor. This factor may apply within the terrain feature as well as in the tactical fringe depending upon the nature of the feature itself.

The effect of a terrain feature on the mobility of a unit, independent of enemy activity, is a function of the trafficability of the feature. To measure this effect we attach to each feature a Terrain Feature Mobility Index (TFMI), between 0 and 1.

We have assumed that units with no terrain features between them detect each other at 15 kilometers. Terrain features normally decrease the range of detection. To indicate the range of detection in a terrain feature we assign to each feature a Visibility Index Radius (VIR) corresponding to the maximum distance within a terrain feature at which one unit can see another. We will take VIR to be a number between 0 and 15.

Certain terrain features provide some cover or some degree of protection from enemy fire. To indicate this protection we associate a Vulnerability Factor (VUFT) with each terrain feature. This factor depends upon the STATE as well as upon the terrain feature involved.

In view of the fact that terrain affects the rate of fire of a weapon, a Fire Suppression Factor-Terrain (FSFT) is a logical refinement. This factor depends upon the STATE of the unit and the particular weapons class under consideration.

4. The Program Including Terrain. The program for the introduction of terrain into the War Game described in PART I has not been completed. However a preliminary analysis of the modifications has been made for a program in BASIC FORTRAN on the IBM 7072 with the boundaries of the terrain features defined by ellipses.

The biggest problem in using the game is entering the large amount of data necessary for a terrain feature. Some method other than manual for obtaining and entering this data would contribute greatly to the introduction of terrain into such war games.

The analysis indicated that the following algorithms must be added to the program of PART I in order to include the terrain:

a. An Algorithm to Determine Whether or not an Element is in a Terrain Feature. The playing area is subdivided into sub-areas (in our case 16 squares). Associated with each sub-area is a list of terrain features whose tactical perimeters contain one or more points of the sub-area. To determine if an element is in a terrain feature, one starts by determining which sub-area contains the element and then checks the element with each feature contained in that sub-area. If it is in a feature, the factors associated with the feature are transferred to a list kept by the element. If it is not inside a feature but is in only one tactical perimeter, the factors associated with the perimeter are transferred to the element. If the element is not inside a feature but is in more than one tactical perimeter, we list the tactical perimeters containing the element and for each factor the extreme value is transferred to a list associated with the element.

We are assuming here that no terrain features intersect but that the tactical perimeters may intersect.

With the element a circle and the terrain feature and its tactical perimeter ellipses, the above computation can involve finding the intersection of a circle and an ellipse. However, we can obtain an acceptable answer by simply asking whether the center (a point) of the element is inside the ellipse, which is a very easy calculation and which will simplify the program and reduce the storage requirements.

b. Detection Algorithm. The purpose of this algorithm is to furnish for each element a list of all enemy elements which it has detected. It is rather long and involved and we will not take the space to describe it in detail. The method consists of scaling down the basic 15 kilometer detection radius by a factor which depends on the amount of the line of sight which is contained in various terrain features.

In other words, we determine how far a unit can see in a given direction (which, of course, depends on the terrain features in that direction) and then see if there is an enemy unit within that distance.

c. Vulnerability Algorithm. The basic attrition per time interval is reduced by a vulnerability factor determined by the STATE of the unit and the terrain features involved.

d. STATE Algorithm. If an element is within the tactical perimeter of a terrain feature and had been in MARCH STATE during the previous time interval, then its STATE is changed to CONTACT. Otherwise, the STATE is determined as in the basic game.

e. Speed Algorithm. If the unit is in a tactical fringe but not in a feature and it sees an enemy unit then the speed of the unit is half the distance to the closest sighted enemy unit per hour multiplied by the worst SDFT. If the unit is within a tactical fringe but not in a terrain feature and does not see an enemy unit then the speed of the unit is road speed multiplied by the worst SDFT. If the unit is within a terrain feature and sees an enemy unit then the speed of the unit is half the distance per hour to the nearest sighted enemy unit multiplied by the worst of the SDFT and the TFMI. If the unit is within a terrain feature and does not see an enemy unit then the speed of the unit is road speed multiplied by the worst SDFT and TFMI.

f. Fire Suppression Algorithm. The basic rate of fire is reduced by a factor determined by the STATE of the unit and the terrain features involved.

Progress to date with the terrain indicates that the structure of the game possesses the generality and the adaptability essential for an effective experimental war game.

Bibliography

1. DEVELOPMENT OF CENTAUR. A COMPUTERIZED WAR GAME. PART I. GENERAL CONSIDERATIONS. Charles R. Roberts. (U.S. Army Strategy and Tactics Analysis Group, Bethesda, Md., Stagten No. 62121, Dec. 1962) U. S. Government Research Reports (Office of Technical Services, Department of Commerce, Washington, D.C.).

2. SECOND WAR GAMING SYMPOSIUM PROCEEDINGS, Washington Operations Research Council, Washington, D. C., 16,17 March 1964.

3. THE CENTAUR WAR GAME - CLINIC. U.S. Army Operations Research Symposium, Proceedings - Part I, Durham, N. C., 1963, pp. 105-128.

4. Lanchester, Frederick William. Aircraft in Warfare. Constable and Co., London, 1916; or D. Appleton and Co., New York, 1917; see Chapters V, VI. Material also available in World of Mathematics, Simon and Schuster, New York, 1956; pp. 2138-2157.

AN ANALYTIC PROCEDURE FOR EVALUATING BLAST-FRAGMENTATION WARHEADS IN AIR DEFENSE

by

Mr. Robert C. Banash

U. S. Army Ballistic Research Laboratories
Aberdeen Proving Ground, Maryland

INTRODUCTION

One measure of the lethality of a surface-to-air missile is the probability (P_K) of the missile warhead inflicting damage on the structure or components of a target according to specified criteria. This probability, known as the single-shot kill probability, is based on the assumption that all components of the weapon system have functioned as designed to deliver the warhead to the vicinity of the target. Currently, most estimates of P_K are obtained from Monte-Carlo sampling procedures (References 1 and 2). In this method a random sample of n engagements is chosen from the distribution expected under tactical conditions. The mean target kill probability, evaluated over the n intercepts is an estimate of the parameter, P_K. The engagements are usually mathematically simulated due to the prohibitively high cost and impracticality of full scale tests. The extensive computations required by these mathematical models make the services of a high speed electronic computer mandatory.

The methods developed in this paper allow the weapon systems analyst, using tables and hand computations, to estimate the P_K of surface-to-air missile warheads which use blast, fragments or both as target kill mechanisms. The intercepting missile is assumed to attack along a trajectory parallel

to the target velocity vector. A similar engagement geometry is frequently encountered by surface-to-air missiles in self-defense roles, e.g., missiles launched from ships under attack and anti-missile missiles launched along inverse trajectories to the attacking missile. The miss-distance* (r) is considered to be distributed as the Rayleigh distribution

$$f(r) = \frac{r^2}{\sigma_g^{\,2}} \exp\left[-\frac{r^2}{2\sigma_g^{\,2}}\right] \qquad r \geq 0$$

Range errors are assumed to be distributed normally with mean located on the target axis and variance, $\sigma_r^{\,2}$.

* Miss-distance is defined, for this paper, as the distance of closest approach between warhead and target centers-of-gravity.

TARGET VULNERABILITY

Blast vulnerability data express target kill as a function of charge weight, altitude and distance of burst from various stations on the target. Based on this information a lethal envelope can be defined about the target. The probability of a blast kill (P_b) is practically 1.0 for a burst occurring within the envelope, and outside the envelope it is 0. For evaluation purposes the blast envelope can be approximated by one or more ellipsoids. The number of ellipsoids used will depend on the accuracy desired. The shape of the envelope will approximate the actual shape of the target for small warheads and grow to a spherical volume for the larger warheads. The ellipsoids considered will be non-intersecting in order that the probabilities of bursting in the individual ellipsoids can be summed to obtain the total chance of blast kill, P_b.

The target will be defeated by fragments if at least one fragment strikes one of the designated vulnerable components (areas) of the target. For simplicity, these areas will be considered grouped into one or more points centered on the target axis.

To an observer on the interceptor, the warhead emits fragments uniformly in a beam bounded by two half-cones of half-angles ϕ_1 and ϕ_2 (See Figure 1). The vertex of each cone is at the warhead c.g., the axis is the missile axis.

To an observer on the target, the fragments would appear to have velocity

$$\vec{V}_f' = \vec{V}_f + \vec{V}_m - \vec{V}_t$$

where:

$\vec{V}_f$ = fragment velocity with respect to the missile (static)

$\vec{V}_m$ = missile velocity

$\vec{V}_t$ = target velocity

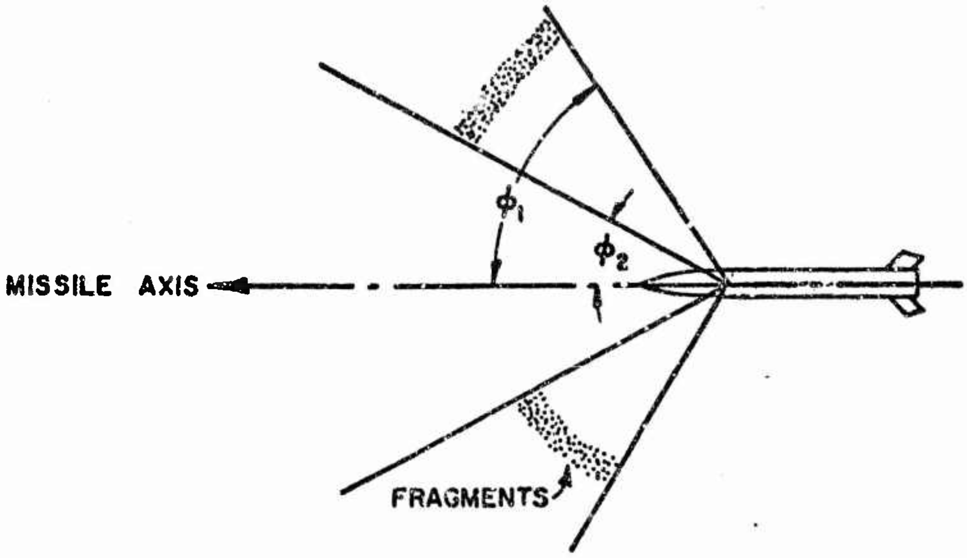
The beam angles, to the observer, would be ϕ_1' and ϕ_2' (Figure 2 and Figure 3) where:

$$\phi_1' = \text{Arctan} \frac{V_f \sin \phi_1}{V_f \cos \phi_1 + V_m + V_t}$$

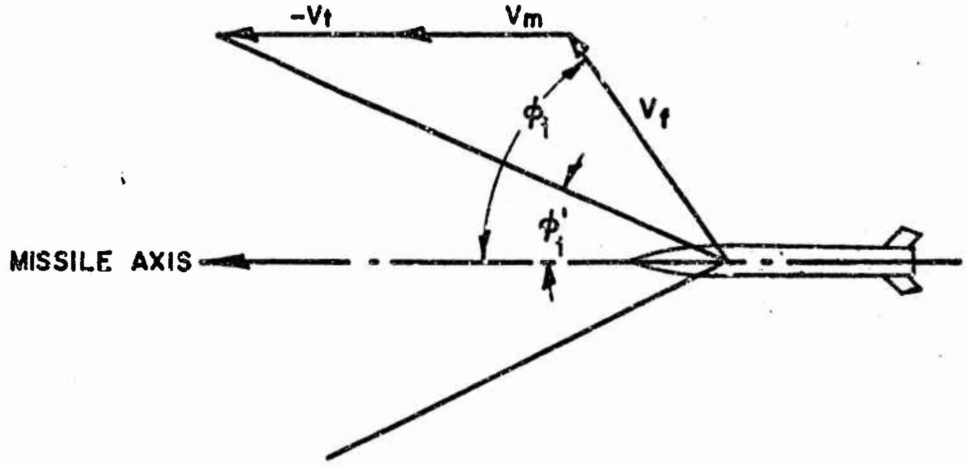
Since V_m is parallel to V_t, a lethal fragment volume can be generated which is bounded by two cones with vertex at the center of target vulnerability, axis coincident with the target axis, and half-angles ϕ_1' and ϕ_2' (Figure 3).

Bursts occurring within the lethal fragment volume would contain the center-of-vulnerability (cv) in the fragment spray. If the cv is in the spray, the probability of at least one fragment striking (and thereby killing) the vulnerable area is

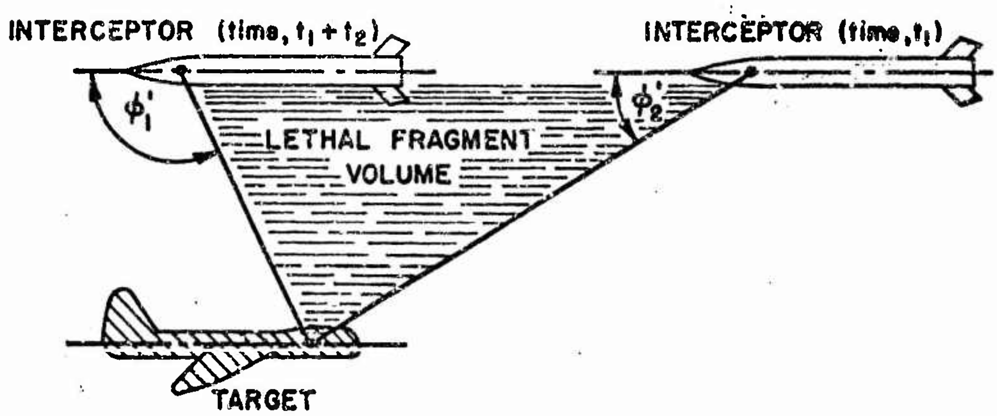
$$P_f = 1 - P\ (\text{zero hits}) = 1 - \exp\left[-\frac{N \cdot A}{d^2}\right] \qquad (1)$$



STATIC FRAGMENT BEAM
FIGURE 1



DYNAMIC FRAGMENT BEAM ANGLE (ϕ_i')
FIGURE 2



LETHAL FRAGMENT VOLUME
FIGURE 3

where: N = fragment density (frag/steradian)

A = vulnerable area

d = distance of fragment travel to target

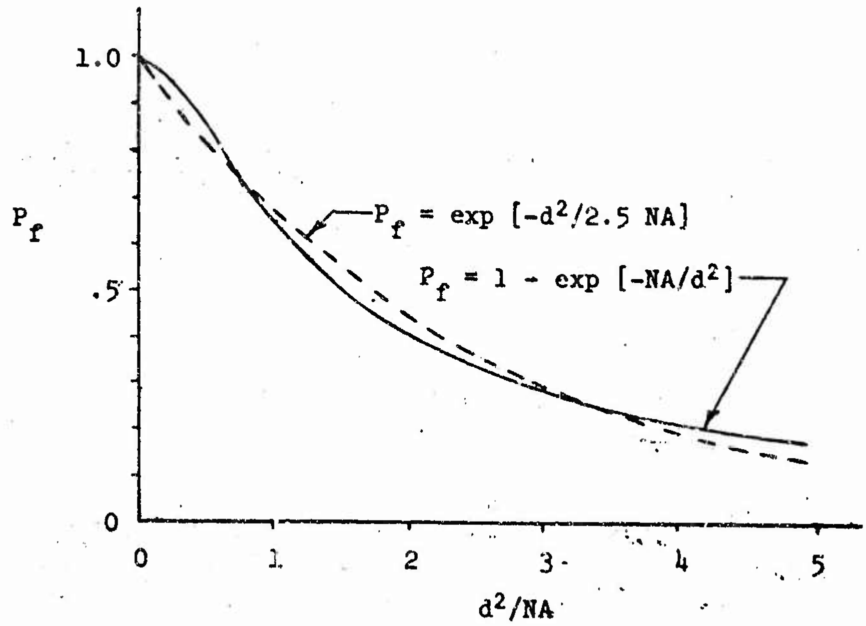
$\exp\left[-\frac{N \cdot A}{d^2}\right]$ = probability of zero hits (Poisson Law)

For computational ease, the Carlton approximation to P_f is used,

$$P_f = \exp\left[-\frac{d^2}{2a^2}\right]. \tag{2}$$

A good fit in most instances is obtained by setting $a^2 = 1.25 \cdot N \cdot A$, see below.

Carlton Approximation to P_f



WARHEAD BURST POINT DISTRIBUTION

A coordinate system is defined about the target as follows: the target flies in the x_1 direction, x_2 is aloft and x_3 to the pilot's right. The origin of the system is located along the target longitudinal axis; the target center-of-gravity (c.g.) will be used in this study. The warhead burst points are assumed to be normally distributed along each of the three axes with variances $\sigma_{x_1}^2$, $\sigma_{x_2}^2$, $\sigma_{x_3}^2$, and mean at the origin or c.g. The distribution of bursts along x_2 and x_3, or errors of guidance, can frequently be assumed equal ($\sigma_{x_2}^2 = \sigma_{x_3}^2 \equiv \sigma_g^2$) or some average taken (Reference 3),

$$\text{e.g.,}\ \sigma_g^2 = 1/2\cdot(\sigma_{x_2}^2 + \sigma_{x_3}^2)\ \text{or}\ \sigma_g^2 = \sqrt{\sigma_{x_2}^2\,\sigma_{x_3}^2}\ .$$

Errors in the x_1 direction will be referred to as range errors ($\sigma_{x_1}^2 \equiv \sigma_r^2$).

The distribution of the miss distance ($r = \sqrt{x_2^2 + x_3^2}$) was valuable in obtaining P_r. This distribution, called the Rayleigh distribution, is easily obtained, as ...

$$\int_{-\infty}^{\infty}\int f(x_2)\, f(x_3)\, dx_2\, dx_3 = \frac{1}{2\pi\sigma_g^2} \int_0^{\infty}\int_0^{2\pi} \exp\left[-\frac{r^2}{2\sigma_g^2}\right] r\,d\theta\,dr$$

$$= \int_0^{\infty} \frac{r}{\sigma_g^2} \exp\left[-\frac{r^2}{2\sigma_g^2}\right] dr$$

$$= \int_0^{\infty} f(r)\, dr$$

EVALUATION

Blast Kills (P_b)

The probability of a blast kill (P_b) is equivalent to the probability of a warhead burst within one of the, k, non-intersecting ellipsoids which define the lethal blast envelope of the target. This may be expressed mathematically for the "j-th" ellipsoid as

$$P_{bj} = \Pr (Q_j \leq 1) = \Pr \left[\sum_{i=1}^{3} \frac{(x_i - b_{ij})^2}{a_{ij}^2} \leq 1 \right].$$

P_{bj} is summed over the k ellipsoids to yield P_b. A method for obtaining P_{bj} was described by Frank E. Grubbs in Reference 4. The needed parts of this article are summarized here.

Setting $U_i = \frac{x_i}{\sigma_{x_i}}$, $A_{ij} = -\frac{b_{ij}}{\sigma_{x_i}}$, $V_{ij} = \frac{\sigma_{x_i}^2}{a_{ij}^2}$ for any one of the ellipsoids, Q, we can write

$$Q_j = \sum_{i=1}^{3} V_{ij} (U_i + A_{ij})^2 .$$

Now U_i is distributed normally with zero mean and unit variance, so that the variate $(U_i + A_{ij})^2$ is distributed as noncentral chi-square (Fisher 1928) with one degree of freedom and noncentrality parameter A_{ij}. Hence, Q_j is the sum of products of constants and noncentral chi-squares. The sum of

weighted noncentral chi-squares may be approximated by fitting the first two moments to the central chi-square (χ^2). The mean and variance of Q_j are found as

$$m_j = E[Q_j] = \Sigma\ V_{ij}(1 + A_{ij}^2)$$

$$v_j = E[Q_j - m_j]^2 = 2\ \Sigma\ V_{ij}^2\ (1 + 2A_{ij}^2).$$

Since the mean and variance of χ^2 are ν and 2ν, where ν is the number of degrees of freedom, then

$\frac{2\ m_j Q_j}{v_j}$ is approximately distributed as χ^2 with $\nu = \frac{2m_j^2}{v_j}$ degrees of freedom. The probability, P_b, is found by summing the probabilities, P_{bj}, over the k - ellipsoids which approximate the lethal-blast-envelope.

$$P_b = \sum_{j=1}^{k} P_{bj} = \sum_{j=1}^{k} P(\chi_{oj}^2, \nu) = \sum_{j=1}^{k} \int_o^{\chi_{oj}^2} f(\chi^2)\ d\chi^2$$

where

$$\chi_{oj}^2 = \frac{2m_j Q_{oj}}{v_j} = \frac{2m_j}{v_j}\ .$$

The Wilson-Hilferty transformation (Reference 5) of χ^2 to an approximate normal variable, t, may provide satisfactory accuracy, where

$$t = [\sqrt[3]{\chi^2 / \nu} - (1 - 2/9\nu)]\ /\ \sqrt{2/9\nu}$$

which in our notation becomes

$$t_{oj} = [\sqrt[3]{1/m_j} - (1 - v_j/9m_j^2)]\ /\ \sqrt{v_j/9m_j^2}.$$

So the probability desired is simply

$$P_r\ [Q_j \leq 1] = P_r\ [t \leq t_{oj}]$$

where t is approximately normally distributed, N[0, 1].

Fragmentation Kills (P_f)

The fragmentation kill probability, P_f, is obtained by integrating the burst probability at each point, weighted by the kill probability at that point, over the lethal fragment volume. This three-dimensional problem can be solved in two dimensions due to symmetry about the target axis. The volume, in x_1, x_2, x_3 space, reduces to the area, V, in the r, y space ($y \equiv x_1$). If a blast envelope is being considered, the area, V, begins outboard of the intersection of the lethal blast envelope with V, see Figure 4, (this intersection can be estimated graphically).

So

$$P_f = \iint_V \exp\left[-\frac{d^2}{2a^2}\right] \cdot \frac{r}{\sigma_g^2} \exp\left[-\frac{r^2}{2\sigma_g^2}\right] \cdot \frac{1}{\sqrt{2\pi\sigma_f^2}} \exp\left[-\frac{y^2}{2\sigma_f^2}\right] dr dy \quad (3)$$

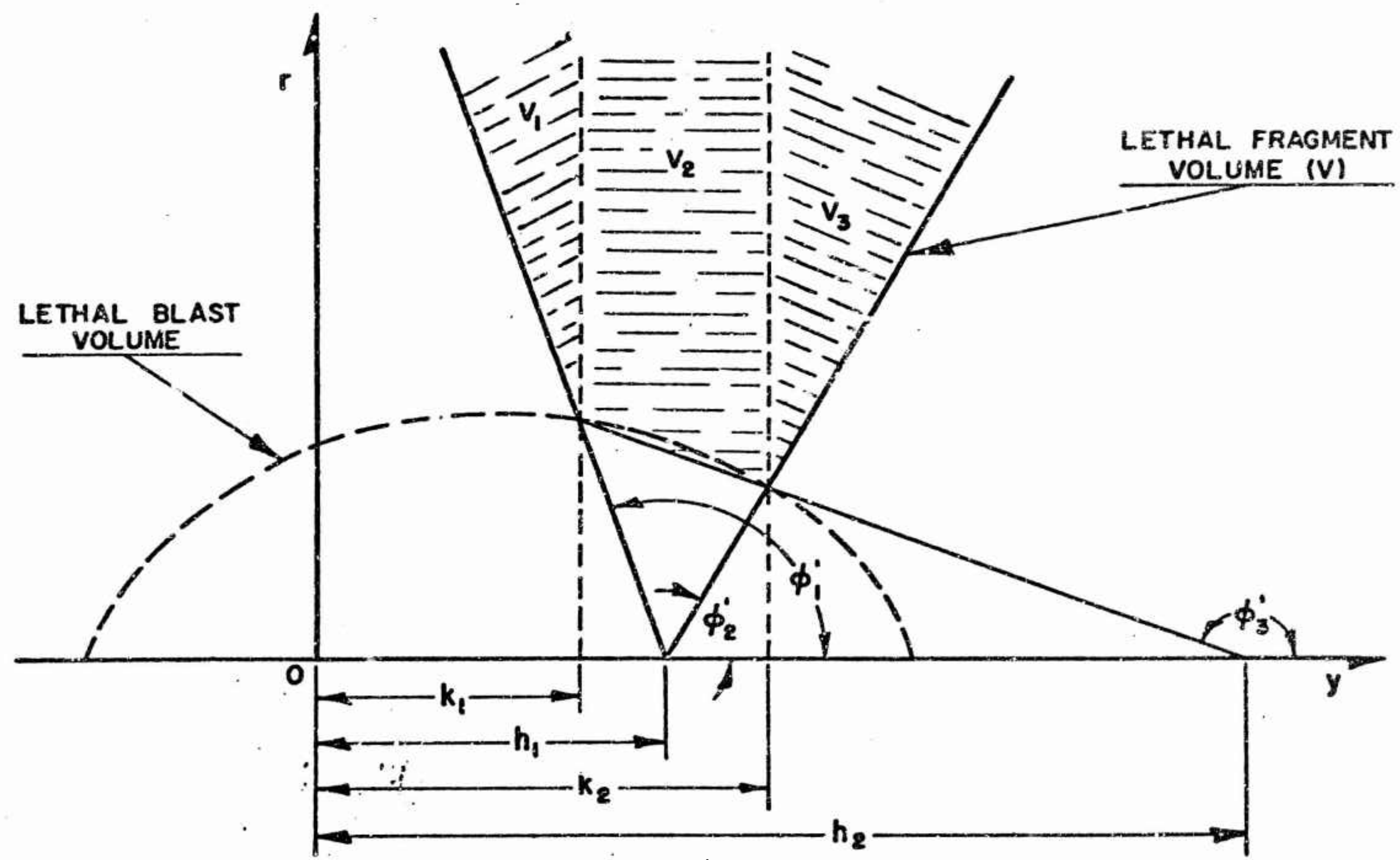
where

$$d^2 = r^2 + (y-h_1)^2.$$

The region, V, can be divided into three regions V_1, V_2 and V_3 to facilitate integration.

$$P_f\ (V) = P_f\ (V_1) + P_f\ (V_2) + P_f\ (V_3)$$

The areas (V_i) are determined by the intersection of the lethal blast volume and the half-cones bounding the lethal fragment volume, (Figure 4). Integrating over the region V_1 gives



REGION OF VULNERABILITY
FIGURE 4

$$P_f(V_1) = \int_{-\infty}^{k_1} \int_{l_1}^{\infty} \exp\left[-\frac{r^2+(y-h_1)^2}{2a^2}\right] \cdot \frac{r}{\sigma_g^2} \exp\left[-\frac{r^2}{2\sigma_g^2}\right]$$

$$\cdot \frac{1}{\sqrt{2\pi\sigma_f^2}} \exp\left[-\frac{y^2}{2\sigma_f^2}\right] dr\,dy,$$

where $l_1 = (y-h_1)\tan\phi_1'$.

$$P_f(V_1) = \int_{-\infty}^{k_1} \frac{1}{\sigma_g^2} \cdot \frac{a^2\sigma_g^2}{a^2+\sigma_g^2} \left(-\exp\left[-\frac{r^2}{2}\cdot\left(\frac{1}{a^2}+\frac{1}{\sigma_g^2}\right)\right]\right)\Bigg|_{l_1}^{\infty}$$

$$\cdot \frac{1}{\sqrt{2\pi\sigma_f^2}} \exp\left[-\frac{(y-h_1)^2}{2a^2} - \frac{y^2}{2\sigma_f^2}\right] dy$$

$$= \frac{a^2}{a^2+\sigma_g^2} \int_{-\infty}^{k_1} \frac{1}{\sqrt{2\pi\sigma_f^2}} \cdot \exp\left[-\frac{(y-h_1)^2\tan^2\phi_1'}{2}\right.$$

$$\left.\cdot\left(\frac{1}{a^2}+\frac{1}{\sigma_g^2}\right) + \frac{(y-h_1)^2}{2a^2} + \frac{y^2}{2\sigma_f^2}\right] dy.$$

Completing the square on y and defining

$$P_\beta = \frac{a^2}{a^2+\sigma_g^2}$$

$$A_i = \frac{\tan^2 \phi_i'}{\sigma_g^2 + P_g} + \frac{1}{a^2} \qquad i=1,2$$

$$B = \frac{1}{\sigma_f^2}$$

$$\mu_i = \frac{A_i h_i}{A_i + B}$$

$$K_i = \frac{h_i^2 A_i B}{A_i + B}$$

$$L(A_i\ \mu_i\ k_i) = \sqrt{A_i + B}\ (k_i - \mu_i)$$

yields

$$P(V_1) = \frac{P_g}{\sigma_f} \int_{-\infty}^{k_1} \frac{1}{\sqrt{2\pi}} \exp\{-1/2[(A_1+B)(y-\mu_1)^2 - K_1]\}\, dy$$

$$= \frac{P_g}{\sigma_f \sqrt{A_1 + B}} \exp\left[-\frac{K_1}{2}\right] \int_{-\infty}^{L(A_1\ \mu_1\ k_1)} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{\omega^2}{2}\right] d\omega \quad (4).$$

$P(V_3)$ is evaluated identically as

$$P(V_3) = \frac{P_g}{\sigma_f \sqrt{A_2 + B}} \exp\left[-\frac{K_2}{2}\right] \int_{L(A_2\ \mu_2\ k_2)}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{\omega^2}{2}\right] d\omega \quad (5).$$

The slope ($\tan \phi_3'$) and the y - intercept (h_2) of the line approximating the boundary between the lethal blast volume and the lethal fragment volume are necessary to perform the integration over V_2. These parameters may be obtained graphically or from the following formulas:

$$\tan \phi_3' = \frac{(k_1 - h_1) \tan \phi_1' - (k_2 - h_1) \tan \phi_2'}{k_1 - k_2}$$

$$h_2 = k_1 - \frac{(k_1 - h_1) \tan \phi_1'}{\tan \phi_3'}$$

$P(V_2)$ is now evaluated:

$$P(V_2) = \int_{k_1}^{k_2} \int_{l_3}^{\infty} \exp \left[- \frac{r^2+(y-h_1)^2}{2a^2}\right] \cdot \frac{r^2}{\sigma_g^2} \exp \left[- \frac{r^2}{2\sigma_g^2}\right]$$

$$\cdot \frac{1}{\sqrt{2\pi}\sigma_f^2} \exp \left[- \frac{y^2}{2\sigma_f^2}\right] drdy$$

$$= \frac{P_g}{\sigma_f} \int_{k_1}^{k_2} \frac{1}{\sqrt{2\pi}} \exp \{- \frac{1}{2} [(y-h_2)^2 \tan^2 \phi_3' (\frac{1}{a^2} + \frac{1}{\sigma_g^2})$$

$$- \frac{(y-h_1)^2}{a^2} - \frac{y^2}{\sigma_f^2}\} dy,$$

where $l_3 = (y-h_2) \tan \phi_3'$. Completing the square on y yields

$$P(V_2) = \frac{P_g}{\sigma_f} \int_{k_1}^{k_2} \frac{1}{\sqrt{2\pi}} \exp \{- \frac{1}{2} [(A_3 + B) (y - \mu_3)^2 - K_3]\} dy$$

$$P(V_2) = \frac{P_g}{\sigma_f \sqrt{A_3+B}} \exp\left[-\frac{K_3}{2}\right] \int_{L(A_3\mu_3 k_1)}^{L(A_3\mu_3 k_2)} \exp\left[-\frac{\omega^2}{2}\right] d\omega \qquad (6)$$

where

$$A_3 = \frac{\tan^2 \phi_3'}{\sigma_g{}^2 \cdot P_g} + \frac{1}{a^2}$$

$$B = \frac{1}{\sigma_f{}^2}$$

$$\mu_3 = \frac{h_1 + h_2 \,(a^2 A_3 - 1)}{a^2 (A_3+B)}$$

$$K_3 = \frac{A_3 \cdot a^2 \,[(h_1-h_2)^2 + B \cdot a^2 \cdot h_2{}^2] + B\, a^2 (h_1{}^2 - h_2{}^2) - (h_1-h_2)^2}{a^4 \,(A_3+B)} .$$

These formulas simplify for certain conditions:

A. if $\sigma_g = \sigma_f \equiv \sigma$

$$P(V_i) = P_g{}^{3/2} (\pm \cos \psi_i) \exp\left[-h^2 \frac{a^2 \sin^2 \phi + \sigma^2}{a^2(a^2+\sigma^2)}\right]$$

$$\int_{L(A_i, \mu_i, k_\mu)} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{\omega^2}{2}\right] d\omega$$

$(i = 1,2)$

B. $h_i = k_j = 0 \; (i,j = 1,2)$

$$P(V) = \frac{P_g}{2\sigma_f} \left(\frac{1}{\sqrt{A_1+B}} + \frac{1}{\sqrt{A_2+B}} \right)$$

C. Both A. and B.

$$P(V) \quad \frac{P_g^{3/2}}{2} (\cos \phi_2' - \cos \phi_1')$$

APPLICATION

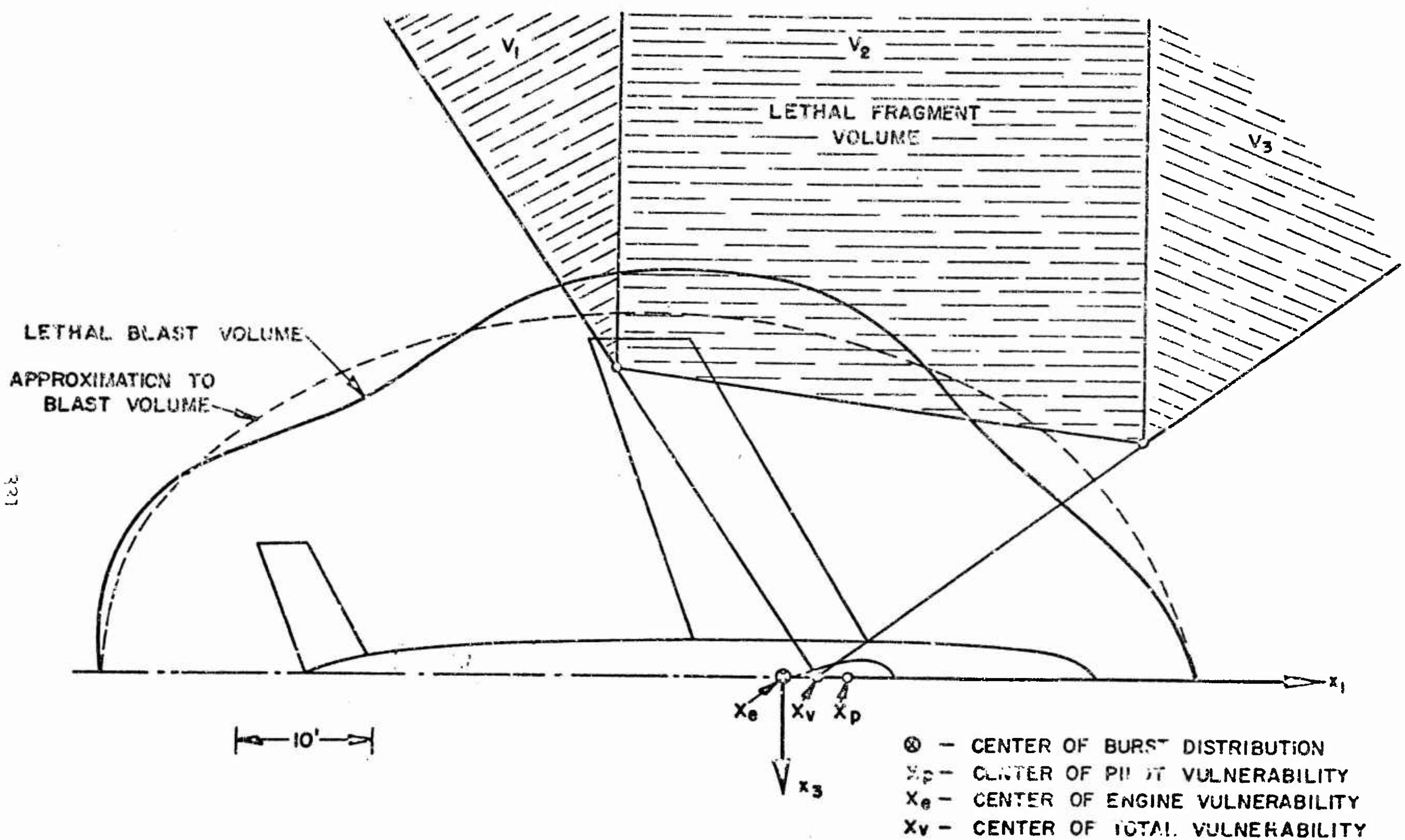
One example is presented to illustrate the use of the derived equations and the resultant probabilities compared with those obtained from the Monte-Carlo simulation described in Reference 2. Additional examples were examined which illustrate the flexibility which can be employed in using this method. These will be presented in greater detail in the publication of this paper as a Ballistic Research Laboratories Report.

The warhead considered for this example consists of 1000 fragments which were ejected at 7000 f/s between the static beam angles of 45° and 135°. The explosive charge of the warhead generated the lethal-blast-envelope illustrated in Figures 5 and 6. The target speed was taken to be 500 f/s and that of the interceptor (V_m) as 1500 f/s. The engagement parameters are:

Engagement Parameters-Dynamic

Beam Angles, ϕ' (deg)	Frag Density, N (frag/ster)	Frag Vel, V_f (f/s)	σ_g (ft)	σ_δ (ft)
36.4-122.9	354	8246	10	15

The vulnerable areas of the pilot, located at (5,0,0), and the engine, at (0,0,0) were summed at the mean location (2.5,0,0) to obtain the analytic estimate of P_K. This approximation was made because the short distance separating components, and the wide fragment spray, would result in both components being jointly in, or out of, the fragment spray with high probability. The boundary between the lethal blast and fragment volumes was a compromise between their projection on the x_1 x_2 - plane (Figure 5) and their projection on the x_1 x_3 - plane (Figure 6). This compromise was necessary to minimize the error arising from the asymmetric blast volume. The values of k_1 and k_2 were obtained graphically.



V1
V2
V3
LETHAL FRAGMENT VOLUME
LETHAL BLAST VOLUME
APPROXIMATION TO BLAST VOLUME
10'
x1
x3
Xe
Xv
Xp
⊗ – CENTER OF BURST DISTRIBUTION
Xp – CENTER OF PILOT VULNERABILITY
Xe – CENTER OF ENGINE VULNERABILITY
Xv – CENTER OF TOTAL VULNERABILITY

FIGURE 5

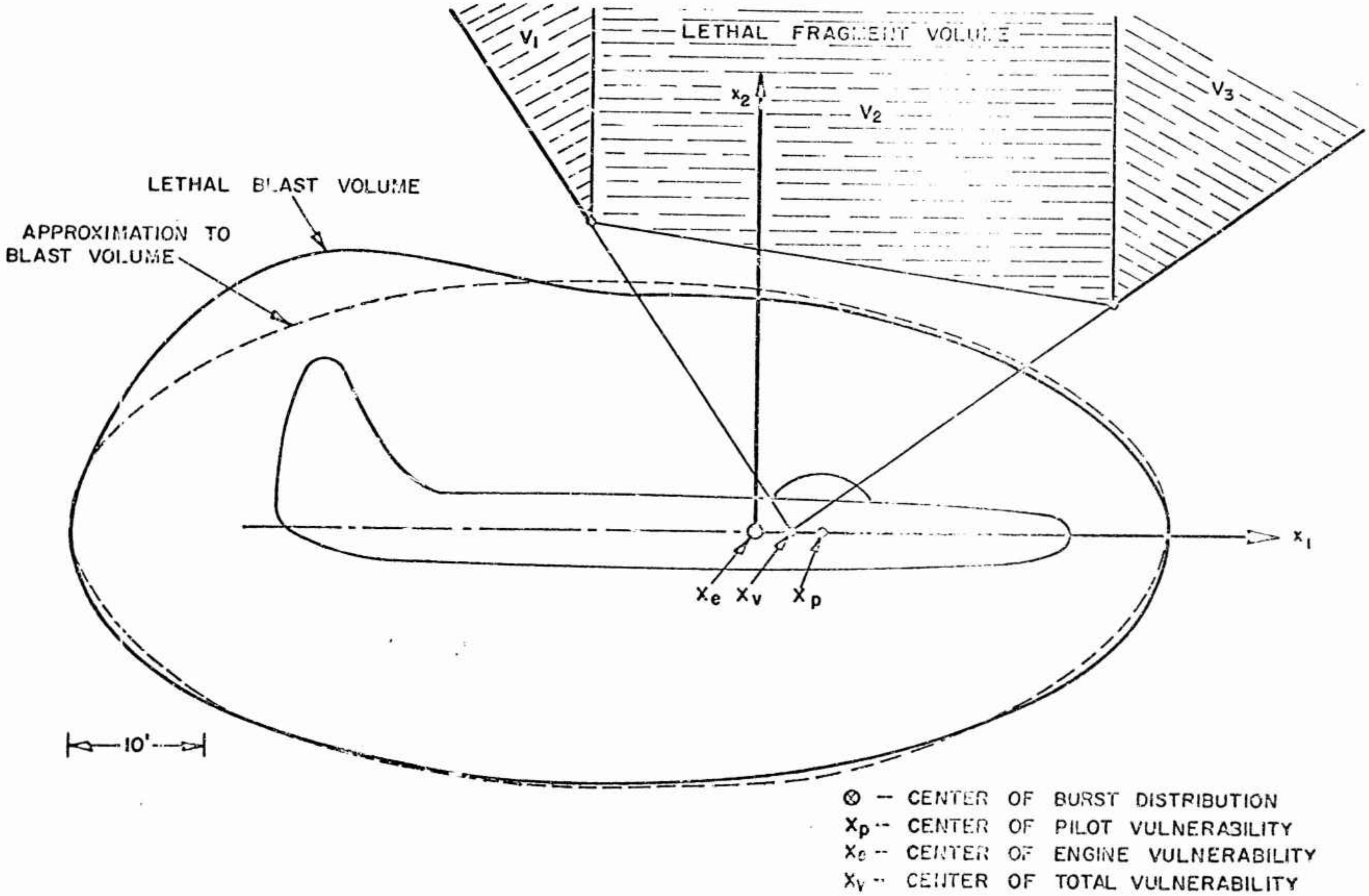

LETHAL FRAGMENT VOLUME
V_1
V_2
V_3
x_2
x_1
LETHAL BLAST VOLUME
APPROXIMATION TO BLAST VOLUME
X_e
X_v
X_p
10'
⊗ -- CENTER OF BURST DISTRIBUTION
X_p -- CENTER OF PILOT VULNERABILITY
X_e -- CENTER OF ENGINE VULNERABILITY
X_v -- CENTER OF TOTAL VULNERABILITY

FIGURE 6

The resultant probabilities obtained by the analytic method are: $P_f = .32$, $P_b = .61$ and $P_K = .93$. Corresponding to these estimates the Monte-Carlo simulation estimated P_f and P_b as .35 and .63 to yield the overall kill probability, $P_K = .98$.

References

1. F. G. King "Lotto Method of Computing Kill Probabilities", Ballistic Research Laboratories Memorandum Report No. 530 (1950).

2. A. D. Stiegler, "A Mathematical Formulation for Ordvac Computation of the Single-Shot Kill Probabilities of a General Missile Versus a General Aircraft", Ballistic Research Laboratories, Memorandum Report No. 1306 (1960).

3. A. D. Groves, "Handbook on the Use of the Bivariate Normal Distribution in Describing Weapon Accuracy", Ballistic Research Laboratories, Memorandum Report No. 1372 (1961).

4. Frank E. Grubbs, "Approximate Circular and Noncircular Offset Probabilities of Hitting", Operations Research 12, 51-62, (1964).

5. E. B. Wilson and M. M. Hilferty, Proc. Nat. Acad. Sci., Washington, D. C. 17,684 (1931).

6. L. W. Campbell and G. A. Beck, "The FORAST Programming Language for ORDVAC and BRLESC (Revised)", Ballistic Research Laboratories Report No. 1273 (1965).

7. Biometrica Tables for Statisticians Vol. 1, E. S. Pearson and H. O. Hartley (eds.) Cambridge University Press, 1954.

8. R. K. Mathur, "A Note on the Wilson-Hilferty Transformation", Bull. Calcutta Stat. Assoc. 10, 102-105 (1961).

THE BIRTH AND DEATH OF A MESSAGE

ABSTRACT

A group of techniques for the analysis of Communications Center and Staff Message Control message-processing operations are presented. The first technique, Automated Flow Process Analysis (AFPA), permits operating personnel to pinpoint ineffective procedures and operations. The second technique, Simplified Message-Processing Simulation (SMPS), permits personnel unskilled in computer techniques to assess system operation under realistic traffic loads. The third, Detailed Message-Processing Simulation (DMPS), permits computer-skilled personnel to perform more refined analysis in major problem areas.

THE BIRTH AND DEATH OF A MESSAGE

Mr. Joel N. Bloom, Manager,
Industrial Engineering Laboratory
Franklin Institute Research Laboratories
Philadelphia, Pennsylvania

Mr. John F. Larison, Technical Director,
Strategic Communications
Department of the Army
Office of the Chief of Communications-Electronics
Washington, D.C.

Project Sponsored by Office of the
Chief of Communications-Electronics

INTRODUCTION

Information is no less a class of materiel than helicopters, tanks, or rifles. Indeed, it is often the prime materiel, for without information, no other materiel can be effectively employed. Information controls the ordering, distribution, use, and destruction of materiel.

Information has a life cycle beginning with birth at a source, continuing through a point of use (destination) and ending with death when it is no longer sufficiently timely to be of use. Military information is often packaged in messages which are transported over wire and/or radio networks from source to destination. The military information manager or communicator must be certain that transportation time does not exceed the information's life cycle, because expired information is about as useful to a commander as expired ammunition is to his troops.

The technological state of present and proposed communciation systems such as AUTOVON and AUTODIN, is such that messages can be rapidly transmitted *between* headquarters. However, *intra*-headquarters message processing has not kept pace with technological advances, and today constitutes by far the major source of delay, particularly during peak periods of crisis and tension. Thus, messages often die (cease to be of use) right in the destination headquarters just a few feet from the men who urgently need the information they contain.

Data from an extensive study of message processing in the U.S. European Command illustrates this point.* On the average, 35 percent of the total handling time is spent processing messages within the

*Cassel, Arne, IMP - Improved Message Processing, Franklin Institute Research Laboratories, Technical Report 1-055, 1965

originating and destination Communications Centers (COMCTRs). As shown in Figure 1, this figure rises to almost 60 percent for flash messages. This is particularly important since, during the period under study, only 20 percent of the flash traffic was within the JCS speed-of-service objective of 10 minutes. The problem is even more severe than the figures indicate, because speed-of-service objectives are defined from receipt in originating COMCTR to ready for delivery in destination COMCTR; thus the figures do not include the very extensive delays in the staff message control units.

Thus, it is clear that headquarters message processing is a necessary and potentially fruitful area for study. The complexity of a typical operation, illustrated in Figure 2 by the portion of the detail flow chart of a typical staff message control unit, underscores the need for readily applicable analytic tools.

Previous programs aimed at studying the headquarters message-processing problem to develop more effective alternatives have tended toward empirical manipulation, that is, on-site modifications in method, procedure, and doctrine. The empirical approach is very inadequate, because it is quite costly and generally cannot be applied in crisis situations where the real problem lies. Analyses by the use of closed-form mathematical models, although attractive in theory, have also failed, because of the difficulty of expressing the complexities of the system in suitable form.

Recent research at the Franklin Institute has led to the development of two techniques for studying headquarters message processing: a static technique termed Automated Flow Process Analysis (AFPA) and two computer simulation techniques, Simplified Message-Processing Simulation (SMPS) and Detailed Message-Processing Simulation (DMPS). This family of analytic techniques has proved quite successful in developing improved alternatives to current practice.

AFPA

Despite the organizational distinction between COMCTR and staff message control (SMC) in the Army and Air Force (they are combined in the Navy), they have been merged for the purpose of this study into a Message Communications Terminal Office (MCTO). The general functions of an MCTO and their interrelationships are shown in Figure 3. These functions are common to message processing in all major headquarters. However, both the specific means for implementing the functions and the effectiveness with which they are conducted vary considerably from headquarters to headquarters and from time to time within a given headquarters.

This variation is partly caused by differences in mission, plant, equipment, and message-traffic type and volume. However, much of the

variation apparently is the result of independent solutions to common problems, rather than a reflection of different local requirements.

Local variations should not prevent standardization of methods and terminology; this standardization is necessary to assure that methods, equipment, layout, and organization are most efficient and to permit personnel transfers to occur without extensive retraining. Also, standardized systems and descriptive terms must be available before system mechanization and automation can be initiated. Automated Flow Process Analysis (AFPA) was developed by FIRL* to provide common descriptive terms and a standardized method of comparing terminal operations.

The AFPA technique permits headquarters communications personnel without flow-charting or systems-analysis experience to document, define in standard terms, and analyze their teletype message-processing methods to determine improvements in these methods. While the technique can be implemented utilizing clerical methods only, punch-card processing of data reduces clerical effort and data-processing time.†

AFPA Technique

The AFPA technique involves analysis of headquarters message processing on the basis of process chains and events. Chains consist of the individual processing steps (events) that are determined by message characteristics (such as precedence and security classification); by the copy or class of copies; and by the operating conditions at the terminal (such as duty hours and minimize). At each terminal, incoming and outgoing messages follow literally hundreds of unique processing chains. However, a relatively small proportion of the chains are used to process most of the total traffic. The following are examples of message characteristics and operating conditions which determine chains:

Incoming, Flash, Secret Message; Advance Copy; Normal Duty Hours

Incoming, Immediate, Confidential Message; Action Copy; Nonduty Hours

Each chain is composed of elemental processing steps which, in AFPA, are called events. These events are carefully defined so that a consistent level of detail is provided for the analysis, and so that terms are used consistently within a given terminal and between terminals.

*Maraist, Paul W. and Barskis, Algird, AFPA - Automated Flow Process Analysis, Franklin Institute Research Laboratories, Technical Report 1-160, 1965.

†However, communications personnel do not have to understand punch-card processing, nor do the punch-card personnel have to understand communications.

Each event is characterized in a number of ways such as time, operator, type of event, and purpose; a dictionary of standard times provides these event descriptors. For example, the type of event is divided into four broad groups: Operations, Inspections, Transportation, and Holds. A portion of the Operation Dictionary is shown in Figure 4.

Event purposes are also divided into four categories: Communication of message, Security, Reproduction and distribution, and Administrative and Quality Control.

A terminal processing operation is defined in terms of a matrix of chains and events. A matrix (shown in Figure 5) lists all events and chains performed and indicates whether or not a given event occurs in a given chain. A separate matrix is made for incoming and outgoing message processing for major terminal message-processing organizations. At a typical headquarters, each matrix may be represented by tens to several hundreds of chains, each of which is composed of ten to fifty (or more) events.

In using AFPA, headquarters message-processing personnel record the characteristics of each event in selected chains (unclassified, low precedence message, normal working hours is a typical starting chain) on specially designed Message Processing Recording (MPR) forms, then code the entries. This material is turned over to EAM personnel for the production of a number of listings for subsequent analysis. A typical listing is the Event-Description Listing which describes each event and the time required to perform it. The listing can be used to locate the events with the longest time cycles; these events offer the greatest potential for reduction in processing time and manpower requirements. Another useful listing is the operator-description listing, which indicates the type and number of operators used to process the messages. The number of different operator functional descriptions is indicative of processing complexity. A large number of different types of operators indicates over-specification; very few types indicate unwieldy and complex work assignments.

Special listings are also produced which may be partitioned by a simple clerical process into detailed flow charts. These flow charts can be utilized to present an overall picture of the processing system. They are valuable in determining the message characteristics, copies, and processing conditions which determine the basic logic of the processing system. Also, the flow charts clearly indicate the relative length of different processing chains and give an effective presentation of processing complexity.

AFPA Results

Illustrative of the type of results which may be obtained from AFPA is an analysis of incoming flash message processing in six major headquarters in the European Command Figure 6 summarizes the events

and reveals the marked differences between headquarters. Figures 7 and 8 show a breakdown of event types. Figure 7 indicates the range of times required to process the advance copy of an incoming flash message from terminal to action agency, command group, or operations center. The times vary from less than 2 minutes at Headquarters F to more than 16 minutes at Headquarters D. The number of individual events (shown by the circled numbers) varies from 8 (for the shortest chain) at Headquarters F to 64 at Headquarters B.

It is instructive to examine the operation and transportation type events separately. (Inspection events are too few to warrent comment.)

The time consumed by operation events varies from 47 seconds to more than 9 minutes. The headquarters with the longest total time (Headquarters D) also has the longest operation time; the headquarters with the shortest total time also has the shortest operation time (Headquarters F).

The percentage of total time devoted to operations (shown on Figure 8) ranged from approximately 15 percent at headquarters A to approximately 60 percent at headquarters B and D; this percentage roughly indicates the division of time between operations and transportations, because inspections consume very small amounts of time.

The apparent cause of these wide time variations is that, in the headquarters with the lowest processing times (Headquarters F), most processing events are eliminated for the processing of the advance copy and are performed subsequently on the copies provided for normal distribution. It should also be noted that Headquarters F was a combined operation performing both communications-center and staff-message-control functions; thus, the message did not require processing by two separate groups.

Transportation times as shown on Figure 7 vary from 60 seconds at Headquarters F to 368 seconds at Headquarters D and represent from approximately 35 to 80 percent of the total processing time. Some of this time difference is based on the physical layout of the headquarters, but it is also partly a function of the type of transportation used between SMC and the COMCTR and between SMC and the action agency, duty officer, or operations center. At Headquarters A through F, pony teletype circuits, pneumatic tubes, and messengers were used for these transfers. At Headquarters A, a flash message is sent over the pony circuit twice in succession to provide SMC an extra copy for advance distribution; mostly for this reason, Headquarters A has a high transportation time, even though it has short operation times (47 seconds). At Headquarters A, SMC and the COMCTR are adjacent so that this high value of transportation time cannot be attributed to physical separation.

At Headquarters F, with the lowest transportation time, a combined COMCTR and SMC eliminates transportation time between these functions.

Overall operations require less average time than transportations; for the six headquarters, operations averaged 9.5 seconds and transportations 33.8 seconds. In part, this great difference occurs because many operations are logical decisions requiring little time to perform; however, the events with the longest times involved transportation of the message.

As a consequence of this analysis, it would appear that to reduce flash processing time, most attention should be given to reducing transportation times. The following three factors are suggestive of what may be done.

> *Physical Layout.* The operations centers, staff message control, and communications center should be close together to reduce transportation time.
>
> *Mode of Transportation.* For short distances, messengers are likely to be the swiftest means of transportation; for medium distances, a pneumatic tube probably is the swiftest means of transportation; for longer distances, a pony teletype circuit is likely to be the swiftest mode of transportation.
>
> *Organization.* Transportation of the advance copy directly from the COMCTR to a predetermined recipient, such as the operations center or staff duty officer, eliminates transportation of the copy from COMCTR to SMC and permits action to be taken before processing the remainder of the copies. Normal SMC functions can be performed on the remaining copies.

Analysis of flash-message processing events by purpose is also useful; Figure 9 shows the categorization of events by purpose.

Most of the time is divided between communication of the message and reproduction and distribution. The variations in proportions of these times are indicative of whether pony circuits are used for communication at the headquarters or distribution is via courier.

Wide variations in two areas do indicate where possible improvements could be made:

1. *Reproduction.* At Headquarters A and B, messages are reproduced to provide additional copies. At Headquarters A, this is accomplished by refeeding the tape through the pony teletypewriter circuits; at Headquarter B, office-type reproduction equipment is utilized to provide additional copies. The need to introduce reproduction time prior to advance-copy distribution requires close examination, since alternate methods can often eliminate reproduction time entirely from the advance-copy processing cycle.

2. *Security.* Approximately 40 percent of Headquarters-D time was devoted to security. This high figure resulted from typing receipts for traffic to the command group; typing receipts on high precedence messages contributed in large part to the 16-minute processing time at Headquarters D.

These examples illustrate the power of AFPA when applied to one processing chain at a headquarters and when used to compare an equivalent chain at several headquarters. When AFPA is applied to a larger number of the chains in a terminal, it becomes an even more powerful analytical tool.

Although batching, input rates, and queues are not considered, queues can be located by observation, and operations that cause queues to develop can be investigated intensively.

Techniques such as line balancing (even division of work assigned to operators) can be employed with confidence to reduce queues, even through the exact effect on queue time cannot be predicted with AFPA.

SIMULATION

Analysis of the message-processing system by AFPA is very useful in pinpointing areas for potential improvement of operations; however, this technique cannot analyze the buildup of queues or the effect of traffic variation on system performance since it is a static analysis tool. To handle these system factors, two computer simulation techniques were developed which not only provide a dynamic picture of the system in operation but also may be readily employed to test the system under differing traffic conditions, tension situations, and changes in processing doctrine and procedure. This testing can occur without disruption of the operation of the system.

SMPS

The first simulation technique, called Simplified Message-Processing Simulation (SMPS),* is an expansion from an AFPA analysis. It is designed to enable communications personnel not experienced in computer programming to evaluate a message-processing system under dynamic conditions.† Communications personnel need only complete detailed worksheets describing system operations; they do not need programming ability.

*Jasoby, Kathe and Fackenthal, Diana, SMPS - Simplified Message-Processing Simulation, Franklin Institute Research Laboratories, Technical Report 1-161, 1965.

† A 7090 computer is required for running the programs.

The need for SMPS arises because communications personnel at a military headquarters must be continually aware of the capabilities and effectiveness of their current message-processing systems, not only with respect to current traffic but also with respect to crisis conditions which may occur. Within the limits of military regulations and command structure, communicators are able to develop changes to improve system operation. However, changes should not be implemented unless there is assurance that the total system operation will be improved; therefore, methods for evaluation of changes prior to implementation are required. Because of the differences in needs, regulations, and traffic at different headquarters, the personnel at the individual headquarters (rather than a higher agency) are best suited to evaluate their own systems. The likelihood that these communication personnel have programming background or inclination is very small.

SMPS has three interdependent parts: a system configuration consisting of message-processing and decision activities; a system input, consisting of the incoming and outgoing message traffic during the time of the simulation run; and the simulation output statistics and related data for analysis.

System Configuration

In SMPS, the system configuration is developed from a series of preprogrammed modules of building blocks* in the following fashion.

Consider the portion of an AFPA flow chart shown in Figure 10. The simplified flow chart shown in Figure 11 is constructed by grouping sequences of AFPA events into uninterrupted tasks so that the personnel and equipment involved would not be interrupted to perform any service for any other message and merging identical chains.

Each task (each block on the simplified flow chart of Figure 11) then is described in terms of personnel and equipment required, next tasks to be performed under what conditions, and processing time expected, either in numbers or as a formula in terms of message characteristics as shown in Figure 12.

The task-definition data is used to match each task to a SMPS module.†

*SMPS is a language derived from the macro-assembler capabilities of IBM's GPSS II. The building blocks of SMPS include a set of GPSS variables defined in terms of the parameters of a GPSS transaction, a set of functions for the generation of GPSS parameters from a deck of cards generated independently to describe a message sample, a few other GPSS system variables, and a set of GPSS macro instructions.

†In unusual cases, it may be necessary to divide a task into several simpler tasks to find a match.

The SMPS modules are divided into six categories. The first category contains general-purpose modules which involve queue number, personnel or equipment identification, next module, and time factors; the time factors are for processing or batching. The second category contains decision modules for usual decisions. The third category contains modules which represent transportation and contain facilities to record transit times within the simulation and for external statistical analysis. The fourth category contains a set of modules to permit delivery at regular intervals. The fifth group contains the modules used to control the flow of messages into the model. The sixth group contains flexible modules which allow most unusual tasks to be performed without requiring knowledge of GPSS.

Figure 13 shows a number of typical modules. The circled notation in the A1 module represents entries corresponding to Task K (tear off message from TT receiver) in the simplified flow chart of Figure 11. In practice, these entries would be made and coded on suitable worksheets where processing times are functions of message characteristics; for example, poking time may be expressed (in seconds) as

$$67 + 4.3 \times (\text{number of addressees}) + 9 \times (\text{number of lines})$$

A series of rules are provided to develop the necessary functions.

System Input

If the goal of the simulation is to improve system operation for current message traffic, then actual samples of a day's or a week's traffic should be used as data for the simulation. If the goal is to improve system operation for different traffic conditions (crisis, for example), it is more satisfactory to use a statistical model to generate a traffic sample.

If current traffic is used, worksheets are prepared from an actual message sample. For each message, the user lists the values of these message characteristics: precedence, classification, number of pages, number of addressees in header, number of copies for local distribution, special security classification, special characteristics, and offline crypto requirements.

The arrival time in minutes, the message identity or in/out number, the total number of lines, and the number of tapes or channels required to transmit the message are computed by a special input program. This program also generates punch cards which contain the specified message characteristics in the form required by the SMPS program; these cards form the input message deck.

If a statistical model is used, worksheets are prepared to describe the statistical properties of the traffic sample. The properties the

user supplies are total number of messages in the sample, frequency of message receipt and transmittal (that is, the number of incoming and outgoing messages per hour according to the days of the week and times of day each frequency is expected to begin), and message characteristics, (as he would for an actual sample).

A different input program then is used to generate the input message deck.

The message-processing system must handle service messages in addition to the normal traffic flow. These messages must also be represented in the simulation, because processing operations for them require personnel and equipment time which would otherwise be spent on normal traffic. However, characteristics for these messages need not be identified, because the loading of equipment and personnel represented by such traffic can be handled by task-type modules.

Output Statistics

The most important question in evaluating a message-processing system is, "How long does it take a message to get through the system?" This information is most meaningful in terms of the cumulative distribution function of the total transit time through the system; however, it may also be important to know the time through major subsystems, as well as the time for messages with special characteristics.

A major output of SMPS is a deck of cards; each card contains all the characteristics of a message, a transit time either through the entire system or through a major portion of the system, and an identifier specifying the meaning of the transit time given. Thus, this card deck can be processed manually, by EAM equipment, or by computer to select the messages with the characteristics of interest and to determine the transit-time distributions for these characteristics.

The other output of SMPS is a printout that includes tables which give the fraction of total number of messages with transit times less than each increment of an accumulating time scale, the queue statistics which indicate where bottlenecks occur (Figure 14), and personnel and equipment utilization tables (Figure 15).

DMPS

IN SMPS, the merging of individual events into uninterrupted tasks obscures some of the system detail. This offsets the relative simplicity and cost of application of SMPS. If the importance or the criticality of a message-processing system dictates analysis at a finer level of detail, use of the Detailed Message Processing Simulation (DMPS) technique is warranted.

DMPS* permits simulation of extremely fine system details, for example the effect of a rules change for batching messages for pneumatic-tube transport during specific hours of the day, or the effect of certain operators pitching in to help overloaded operators. Application of DMPS requires personnel skilled in both communications and computer programming languages, and is therefore not suitable for field use. However, it is appropriate for use by agencies concerned with overall communications-system operation or direction to assess the precise effects of major changes in doctrine or procedure (for example, change from 4 to 3 precedence levels).

DMPS employs IBM's General Purpose Systems Simulator II (GPSS II) language.† Programs in this language can be compiled to run on the IBM 7090, 7094, 7040, and 7044 computers. A detailed discussion of the application of the DMPS technique is beyond the scope of this paper.

Typical Simulation Results

A typical application of simulation is to assess the effect of crisis traffic on a system configuration which operates satisfactorily on normal traffic. Data from a number of previous crises were used to construct a message sample that reflected the increases in volume, level of classification, and precedence which a crisis causes. Figure 16 compares the mean transit times through the same configuration for the normal and crisis message samples. Increases range from a doubling of time for flash traffic to an almost six-fold increase for routine traffic. Using the queue statistics data, it is possible to pinpoint the specific areas that contribute most heavily to increased transit times. Figure 17 shows the fractional change in the critical queues. With this data, it is possible to develop initial alternatives; the dark bars on the figure are areas of particular interest.

For instance, SMC routing can be streamlined and improved by making greater use of routing keys such as attention lines, originator office symbols, or subject lines or codes in the message. These keys would permit a large portion of incoming traffic to be routed by clerical or automated means instead of by highly trained routers. Similarly, reference-message look up can be limited to those determined to be essential by the router.

Receipting and logging can be considerably simplified and improved by elimination of unnecessary paper work and by standardizing forms, files, and signature systems.

*Greene, Sylvan and Perry, Patricia, DMPS - Detailed Message-Processing Simulation, Franklin Institute Research Laboratories, Technical Report I-162, 1965.

†General Purpose Systems Simulator II, International Business Machines Corporation, Form B20-6346-1, 1963.

Another important bottleneck is reproduction of incoming and outgoing messages. Investigation reveals that an excessive number of copies are often reproduced and routed to agencies that have no interest in the subject matter. To reduce reproduction processing time, agency distribution and number of copies furnished should be reduced to a minimum based on need-to-know.

In some headquarters, pony circuits are used to transfer messages from one area to another. While it is desirable to eliminate manual handling, pony-circuit transfer between closely adjacent areas actually tends to increase overall handling time. Approximately 3-1/2 minutes are required to relay a 200-word group message over 60-word-per-minute TTY equipment. In addition, bottlenecks frequently occur during peak loads and crises because of equipment limitations. It is usually faster to transfer such messages by messenger, conveyor, tube, or facsimile.

The multiplication of tapes for transmission takes as long as or longer than the time required to transmit the message. At the present time, multiple tapes are first prepared and then transmitted over the appropriate outgoing circuits. Combining these two sequential functions into a single operation can reduce message processing time considerably. Equipment is currently available which permits message headers to be typed at a central keyboard and transmitted serially on line to each transmitting station. The body of the message then is transmitted simultaneously to all circuits.

Both bottlenecks and the number of misrouted and missent messages can be reduced consideralby by simplifying the systems for assigning and updating routing indicators. Such a system might include assigning a permanent routing indicator to each communications center, arranging the Joint Routing Indicator Book by both unit designation and geographical location, and selective dissemination of routing changes.

SUMMARY

This paper described and discussed the use of a graded group of techniques for the analysis of headquarters message processing: AFPA for static analysis by operating personnel; SMPS for dynamic analysis and test of alternatives by operating personnel with access to a digital computer; and DMPS for use by higher echelon computer-skilled personnel on major problem areas. Use of these techniques on a continuing basis for the improvement of terminal and message-control operations can provide significant increases in communications system responsiveness.

However, it is important to sound the klaxon of alarm! Even with improvements in network and COMCTR and SMC message processing, the major bottlenecks in communications processing remain untouched. A communication does not originate in or end at a counter in a message-processing unit. It begins at an action officer's desk and is only of value when it

reaches his counterpart in a distant headquarters. Figure 18 shows that the internal headquarters staff consumes approximately seven times more time than COMCTR and SMC operations for outgoing high-precedence traffic. Similar relationships are found in incoming traffic. It is in this area that the challenge remains.

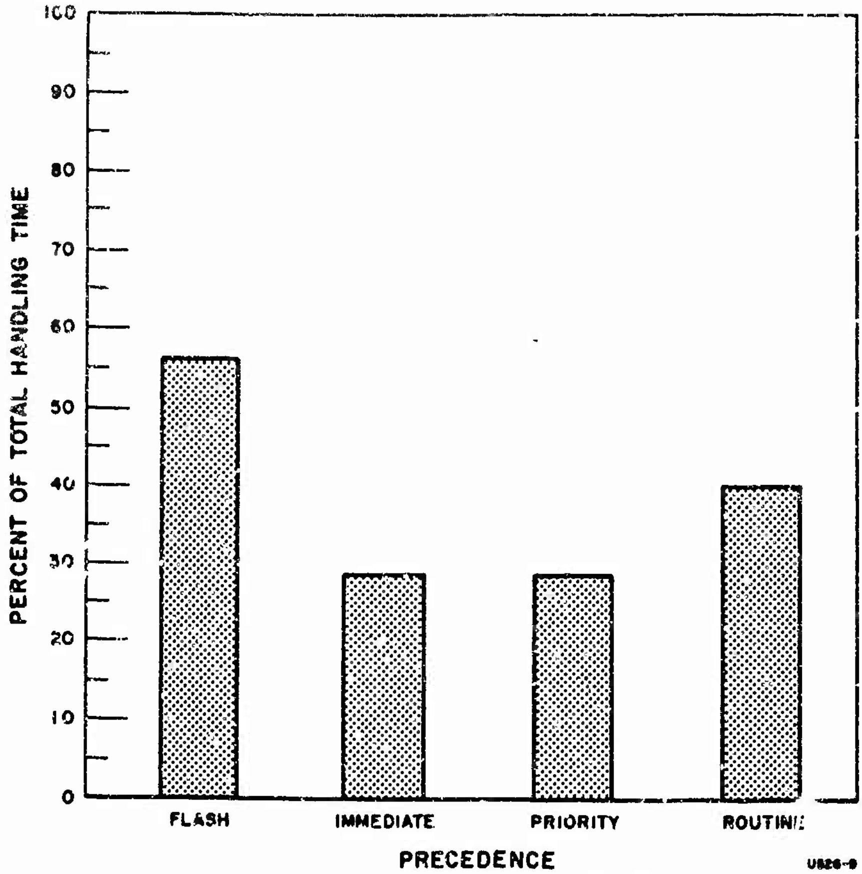


Figure 1. Average Percent of COMCTR Handling Time to Total Handling Time

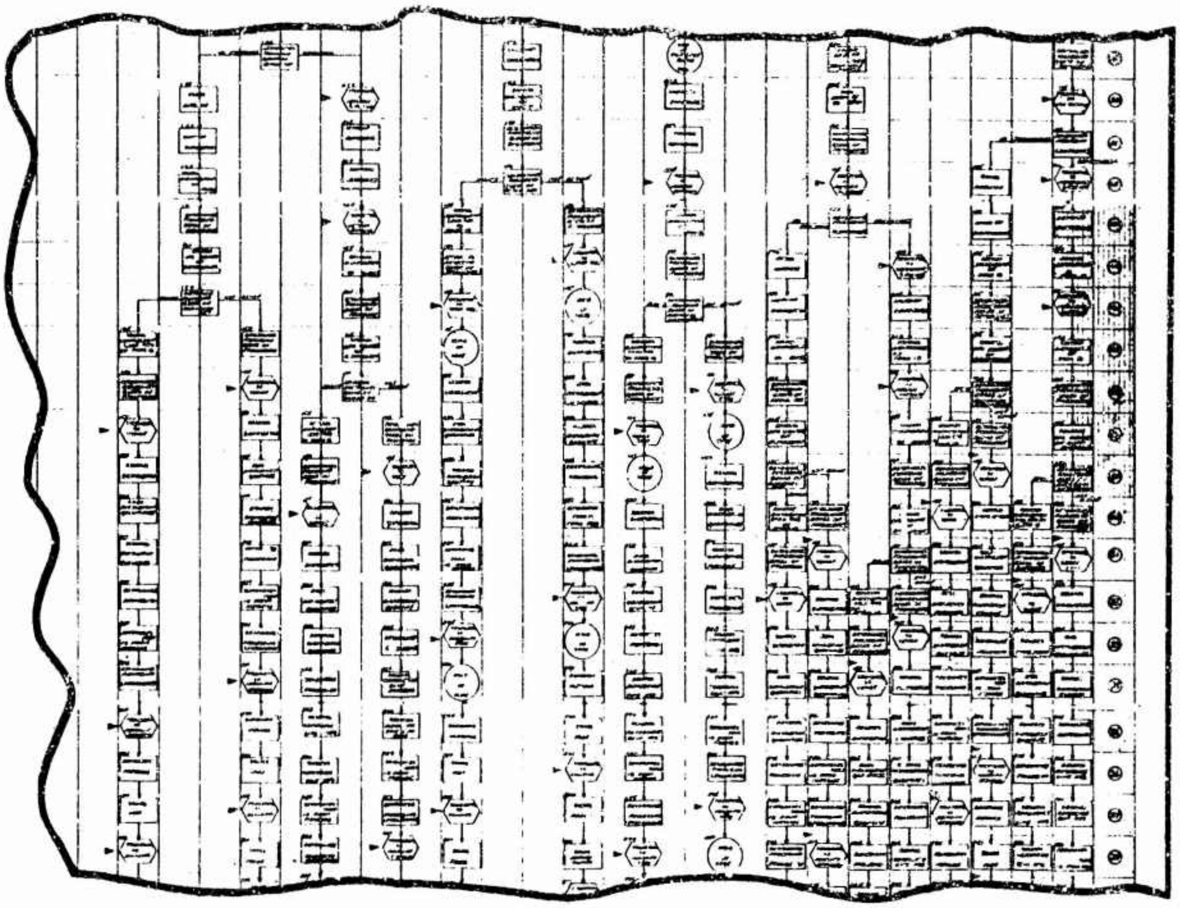


Figure 2. Portion of Detail Flow Chart—SMC Incoming Message Processing

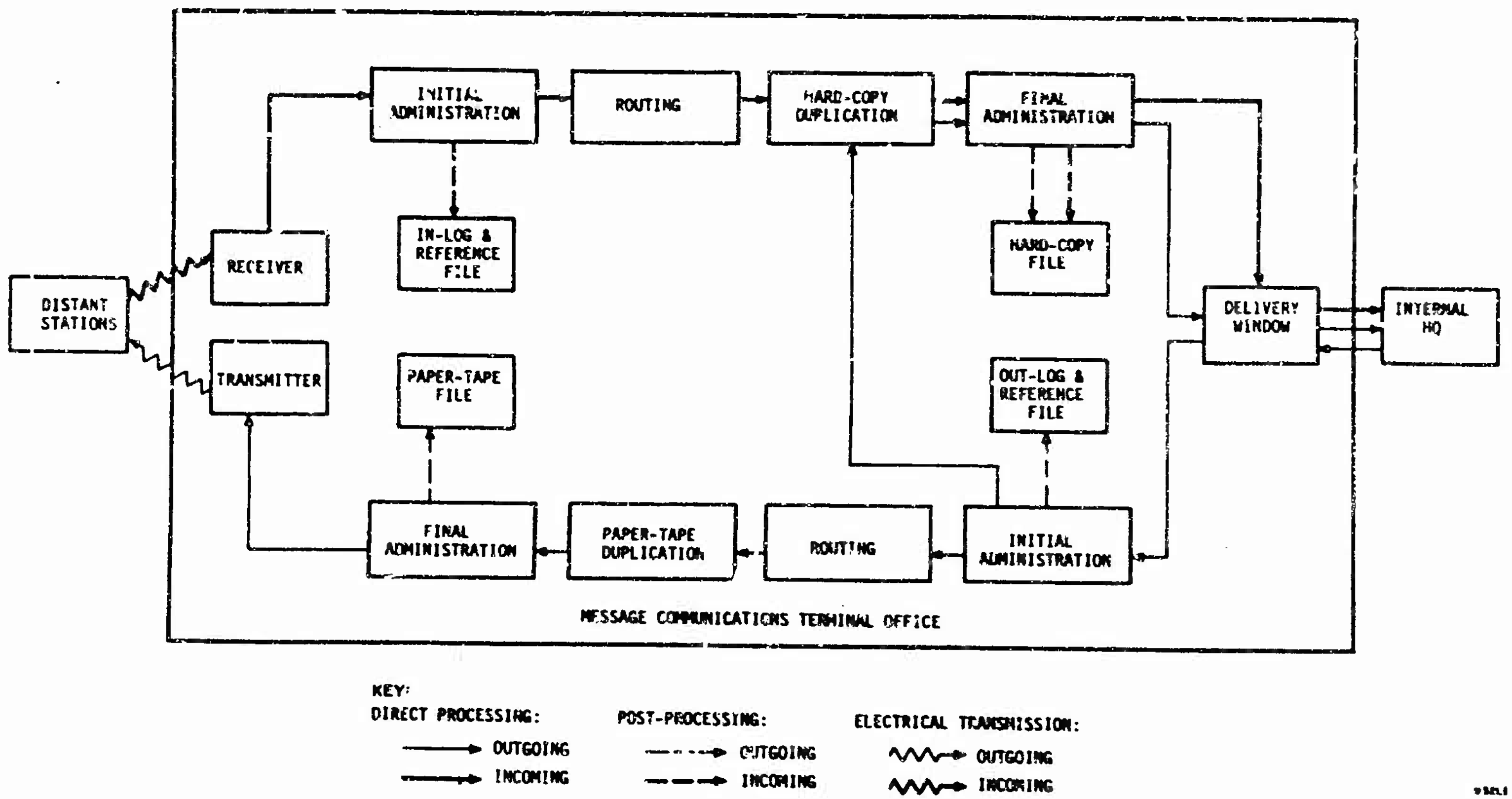


Figure 3. Major Processing Steps in Message Communications Terminal Office

OPERATION

No Physical Change to Message or Other Forms and
Not Machine Operating Function

Logical or Mathematical

0111 Counts
0112 Determines Processing

Data Gathering (Not Verbal)

0121 Looks-Up
0122 Reads
0123 Searches

Verbal Communication

0131 Inqu[illegible]
0132 Notifies
0133 Requests

Sorting-Arranging

0141 Chooses
0142 Collates
0143 Hangs (on rack, etc.)
0144 Sorts

Resulting in Physical Changes to Message or
Other Forms

Transfer Information from Message

0211 Records from Message
0212 Pokes (Cuts Tape)

Figure 4. Portion of Operation Dictionary

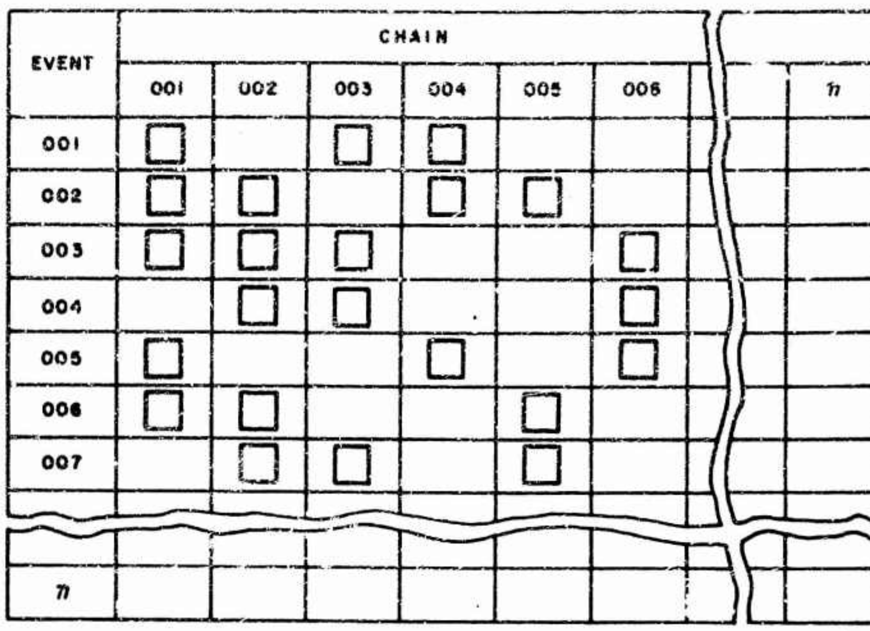


Figure 5. Chain-Event Matrix

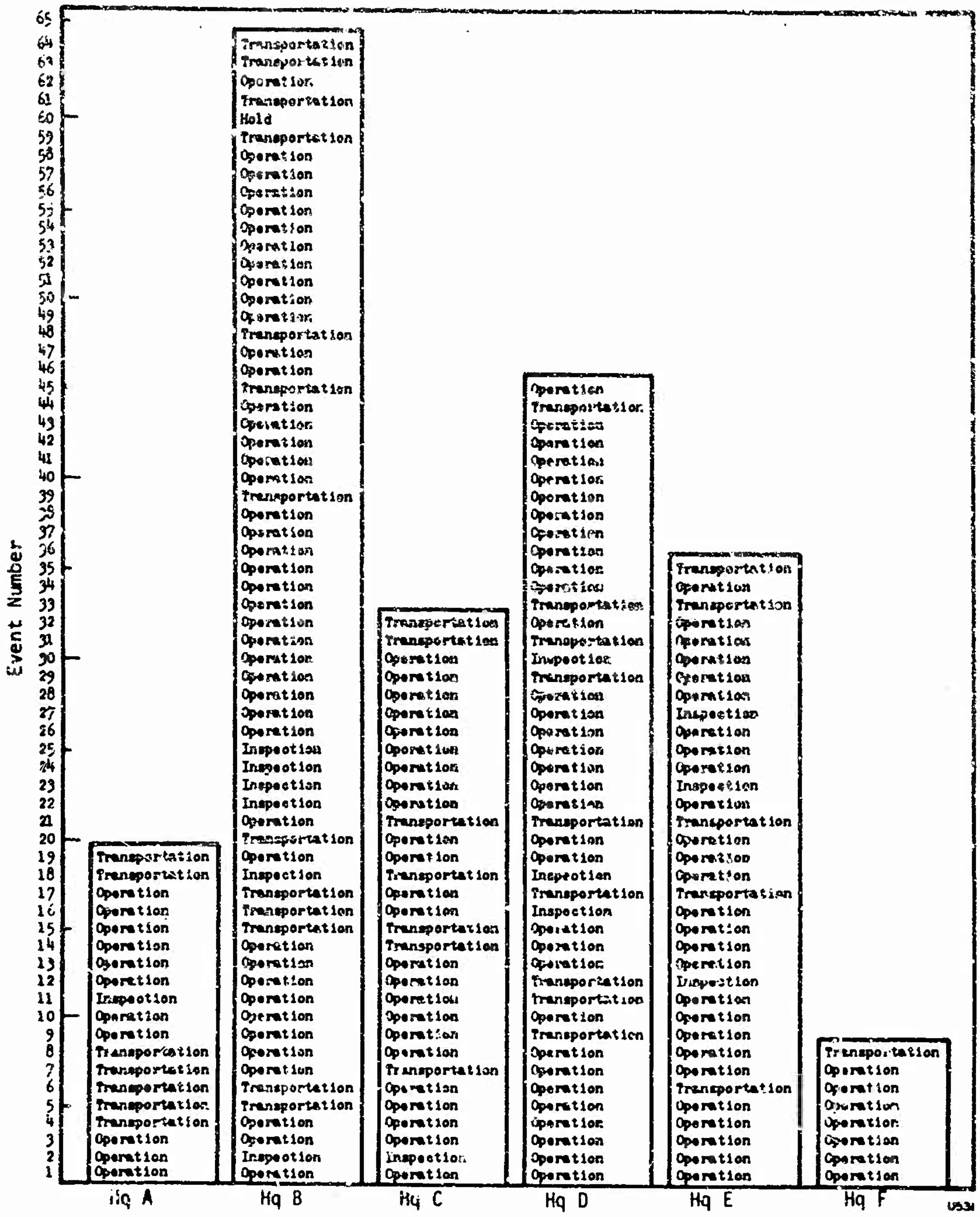


Figure 6. Summary of Events for Incoming Flash-Message Processing, Headquarters A through F

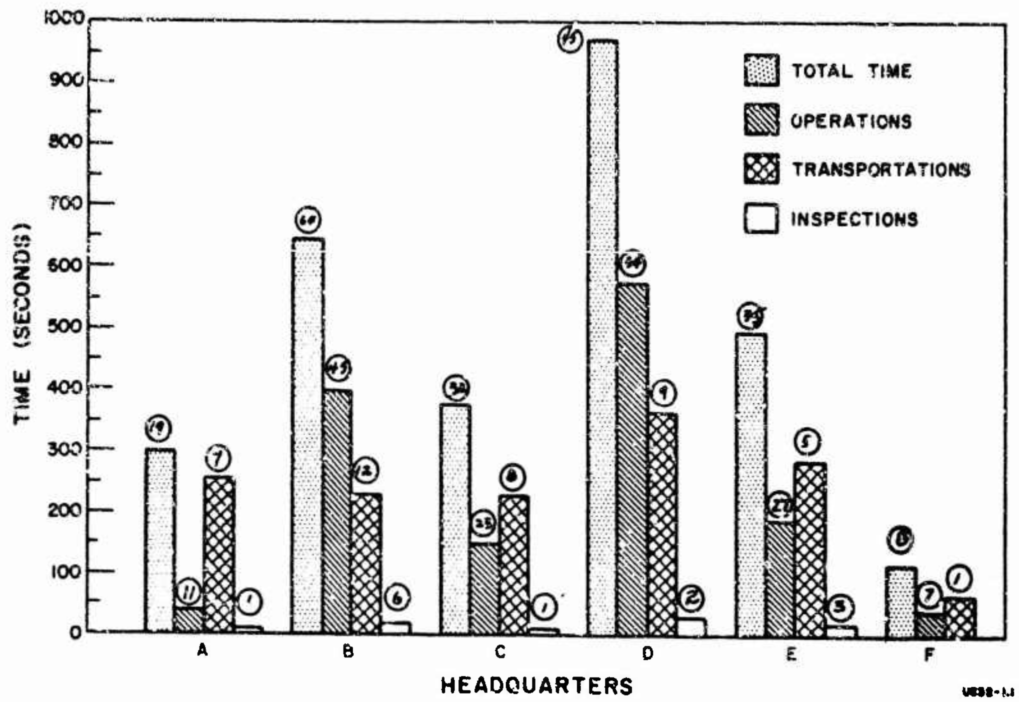


Figure 7. Flash-Message Processing

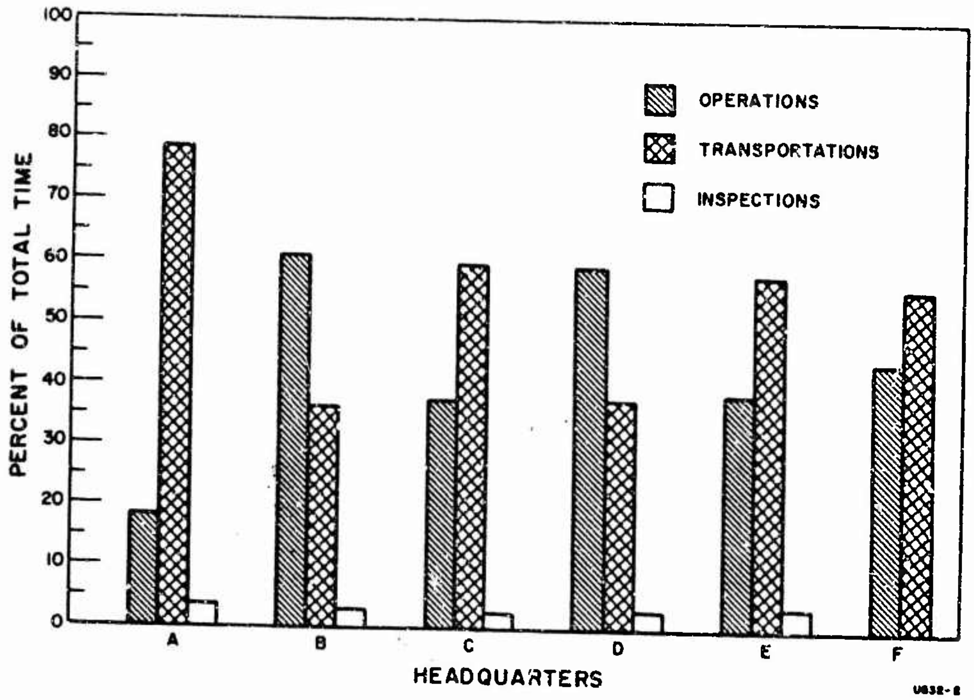


Figure 8. Flash-Message Processing, Percent Total Time by Event/Type

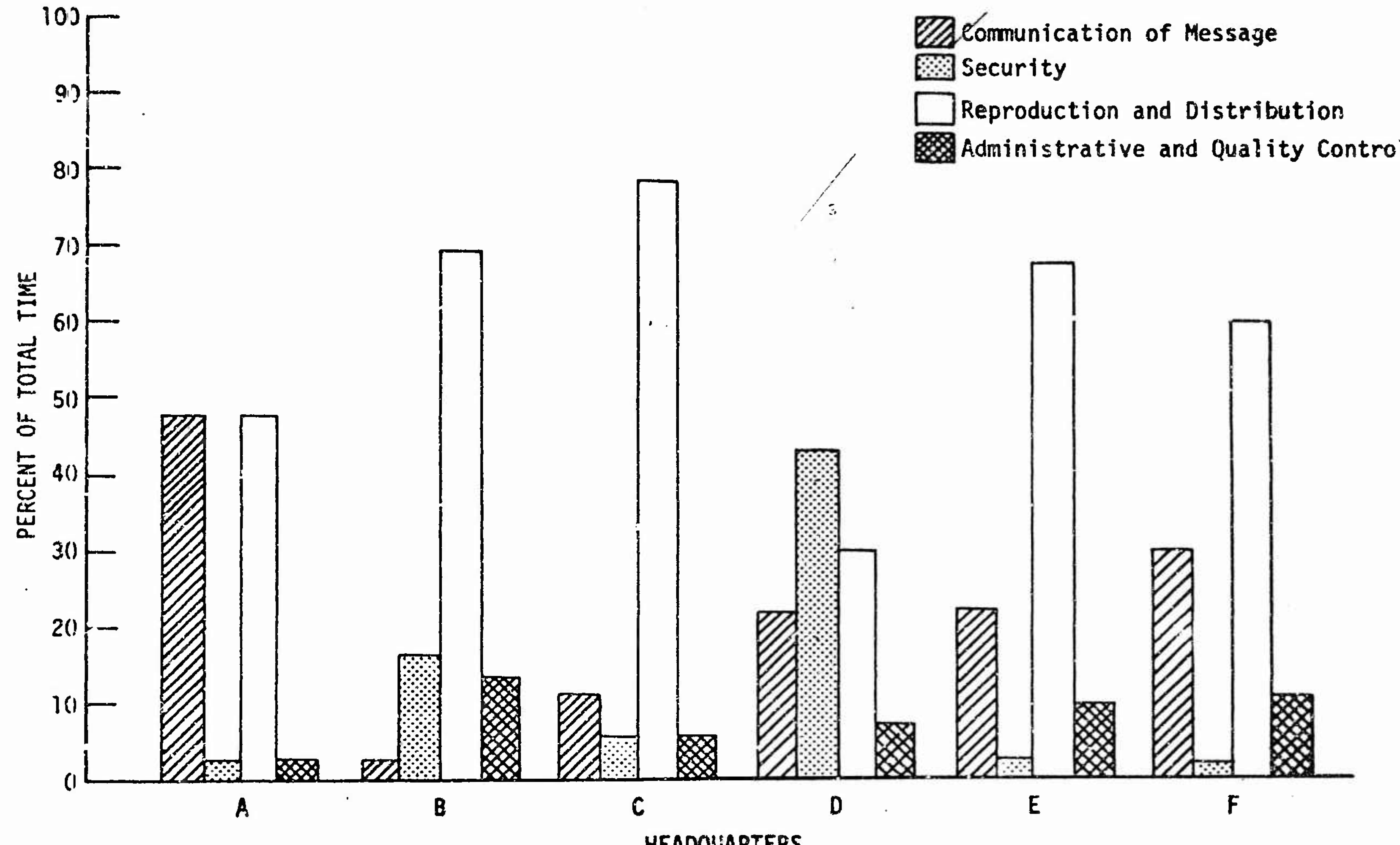


Figure 9. Flash Message Processing, Classification of Events by Purpose

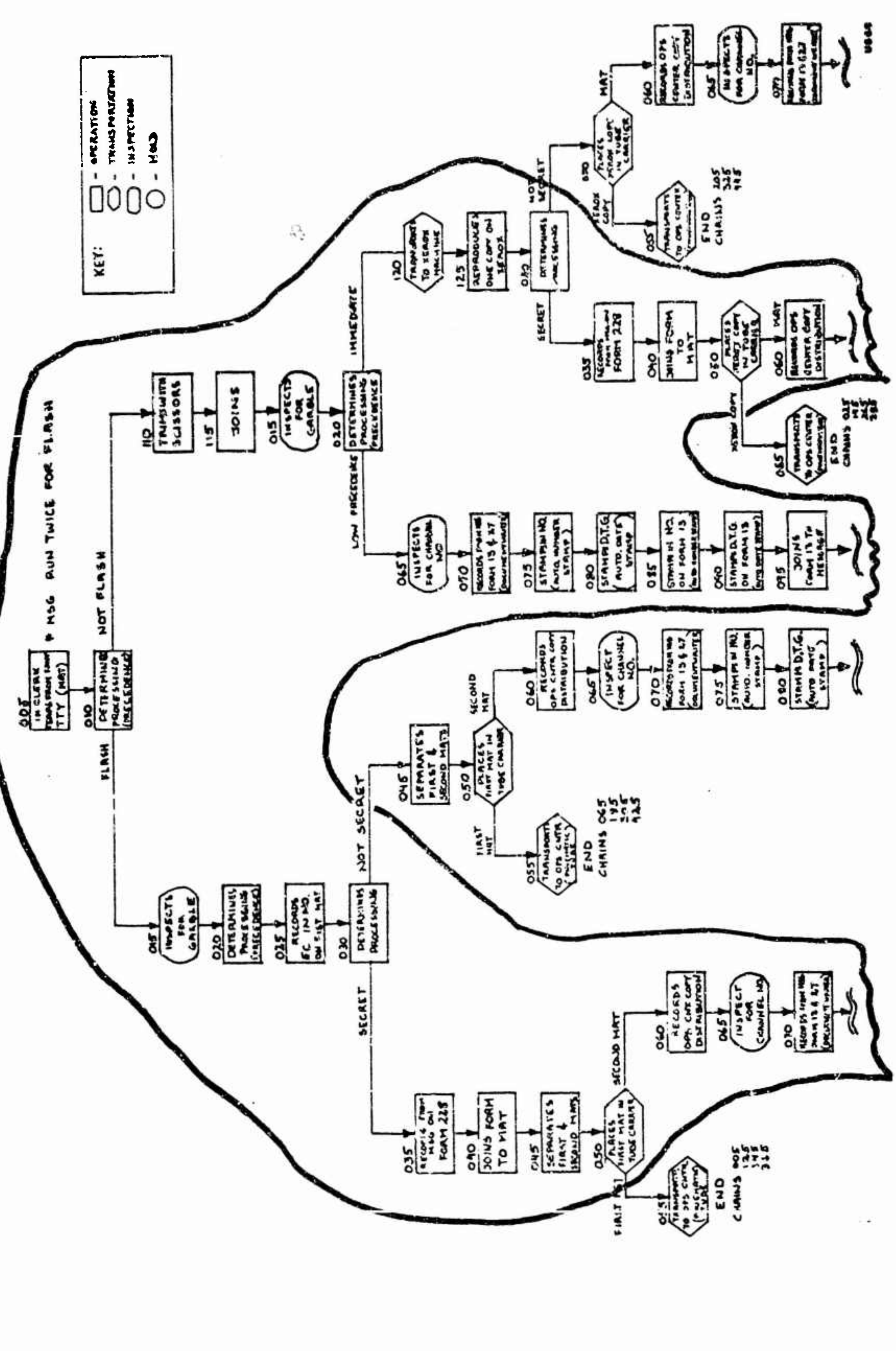


Figure 10. Fragment of an AFPA Flow Chart

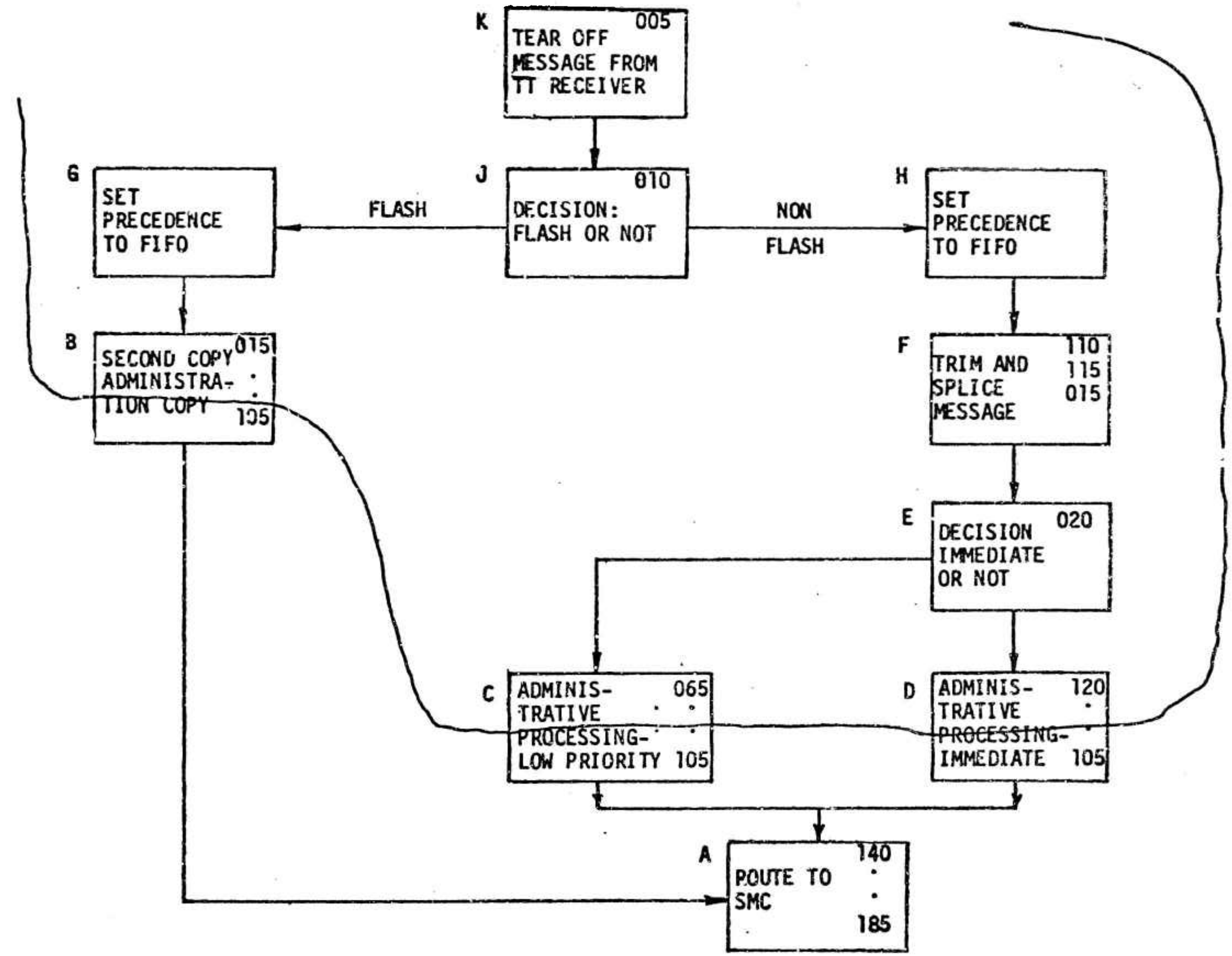


Figure 11. Simplified Flow-Chart Fragment

Headquarters: A | Staff Agency: SMC | Incoming ✓ | Outgoing

Task ID Code	Task Description	Personnel and Equipment Needed	Processing Time		Next Task	Queue Disci-pline	Output Batched
			AFPA Event	Sec.			
K	Tear off message from TT Receiver	SMC Inclerk	005	33	J	FIFO	No
J	Decision flash or not	SMC Inclerk	010	1	GG if flash H if not flash	FIFO	No
G	Set queue discipline to Prec FIFO	-	-	0	B	None	-
B	Administrative processing flash message	SMC Inclerk	015...105	41	A	Prec FIFO	No
A	Routing of Messages SMC	Message Controller	140...185	109	Off chart	Prec FIFO	No
H	Set queue discipline to Prec FIFO	-	-	0	F	None	-
F	Trim and splice message	SMC Inclerk	110...015	30	E	Prec FIFO	No
E	Decision immediate or not	SMC Inclerk	020	1	D if immediate C if not	Prec FIFO	No
D	Administrative processing immediate message	SMC Inclerk Xerox copier	120...105	89	A	Prec FIFO	No
C	Administrative processing low precedence message	SMC Inclerk	065...105	24	A	Prec FIFO	Yes

Figure 12. Task Definition

Name	No. of Variables	Module Description	Meaning of Variables							
			1	2	3	4	5	6	7	8
A1	5	<u>General</u> One individual or piece of equipment takes a message from a numbered queue and processes it for a time specified by two time factors which describe either a rectangular distribution of time or specifies a formula for time. The message is then passed to the next task.	Queue No. 1	Identity of equipment or personnel SMC INCLERK	Next Task TASK J	Time Factor 1 MEAN 33 SEC.	Time Factor 2 SPREAD 3 SEC.			
B1	7	<u>General</u> Same as A1 except two additional time factors are available for batching delay.	Queue No.	Identity of equipment or personnel	Next Task	Time Factor 1	Time Factor 2	Time Factor 3	Time Factor 4	
B2	8	<u>General</u> Same as B1 except that additional equipment or personnel is required.	Queue No.	Identity of first equipment or personnel	Identity of second equipment or personnel	Next Task	Time Factor 1	Time Factor 2	Time Factor 3	Time Factor 4

Figure 13. Typical Modules

QUEUE NR	MAXIMUM CONTENTS	AVERAGE CONTENTS	TOTAL ENTRIES	ZERO ENTRIES	PER CENT ZEROS	AVERAGE TIME/TRANS	TABLE NUMBER	CURRENT CONTENTS
1	4	.01	261	254	97.3	2.52	0	0
2	1	.00	261	261	100.0	.00	0	0
3	1	.00	3	3	100.0	.00	0	0
4	2	.03	33	24	72.7	163.82	0	0
5	1	.00	256	255	99.6	.06	0	0
6	1	.10	5	3	60.0	2213.80	0	1
7	5	.04	259	220	84.9	19.68	0	0
8	1	.00	258	258	100.0	.00	0	0
9	1	.00	3	3	100.0	.00	0	0
10	1	.00	255	255	100.0	.00	0	0
11	1	.00	255	255	100.0	.00	0	0
12	1	.00	9	9	100.0	.00	0	0
13	1	.00	255	255	100.0	.00	0	0
14	3	.06	258	196	76.0	28.76	0	0
15	1	.00	257	250	97.3	.54	0	0
16	3	.14	333	253	76.0	48.76	0	0
17	1	.02	333	291	87.4	6.29	0	0
18	2	.06	330	241	73.0	20.98	0	0
[illegible]	[illegible]	[illegible]	[illegible]	34	[illegible]	50.45	[illegible]	[illegible]

Figure 14. Queue Statistics

STORAGE NR	CAPACITY	AVERAGE CONTENTS	AVERAGE UTILIZATION	NUMBER ENTRIES	AVERAGE TIME/TRANS	CURRENT CONTENTS
2	2	.33	.1665	1100	34.81	0
3	1	.26	.2565	41	719.39	0
4	1	.15	.1532	144	122.33	0
5	1	.41	.4122	8	5924.25	1
6	4	.35	.0884	259	156.97	0
7	1	.39	.3939	188	240.88	0
8	5	.61	.1215	341	204.91	1
9	1	.26	.2567	1294	22.81	1
10	1	.14	.1442	228	72.72	0
11	2	1.17	.5869	333	405.27	0
12	1	.04	.0369	85	43.38	0
13	1	.26	.2630	333	90.82	1
15	1	.20	.1981	257	88.64	0
16	1	.38	.3830	258	170.70	1
17	1	.45	.4465	406	126.4	0

Figure 15. Utilization of Personnel and Equipment

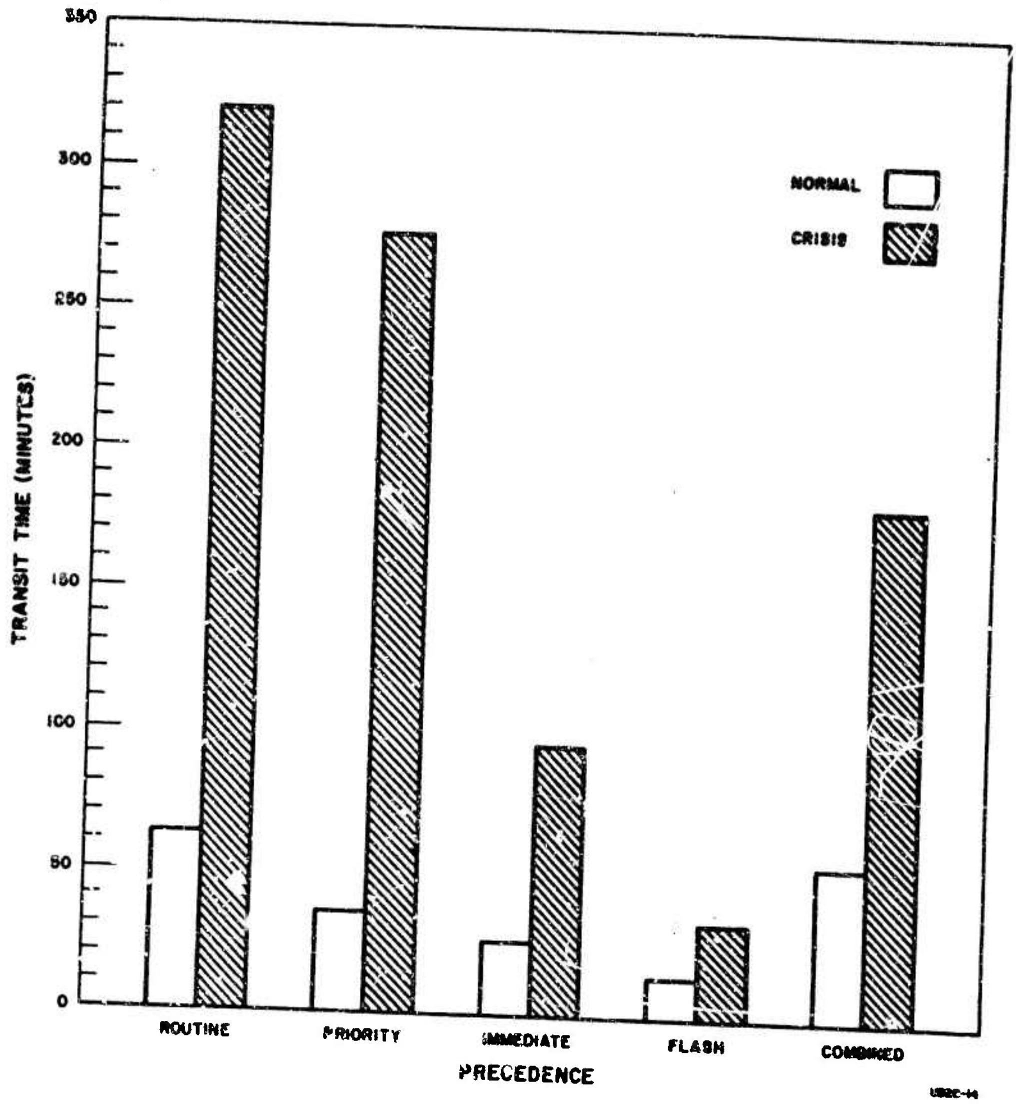


Figure 16. Comparison of Mean Transit Times for Normal and Crisis Runs, Incoming Traffic

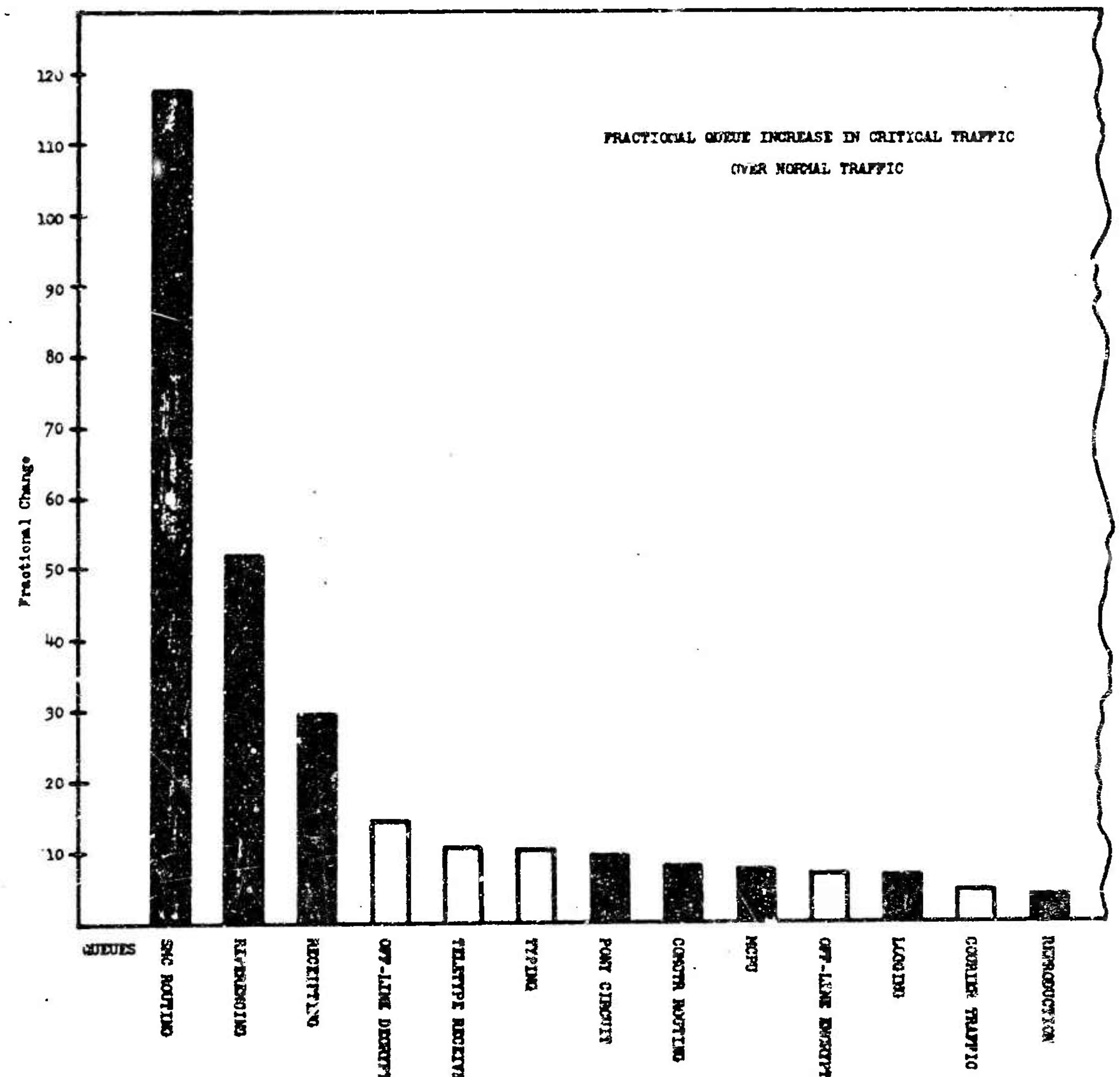


Figure 17. Critical Queues in Crisis Run

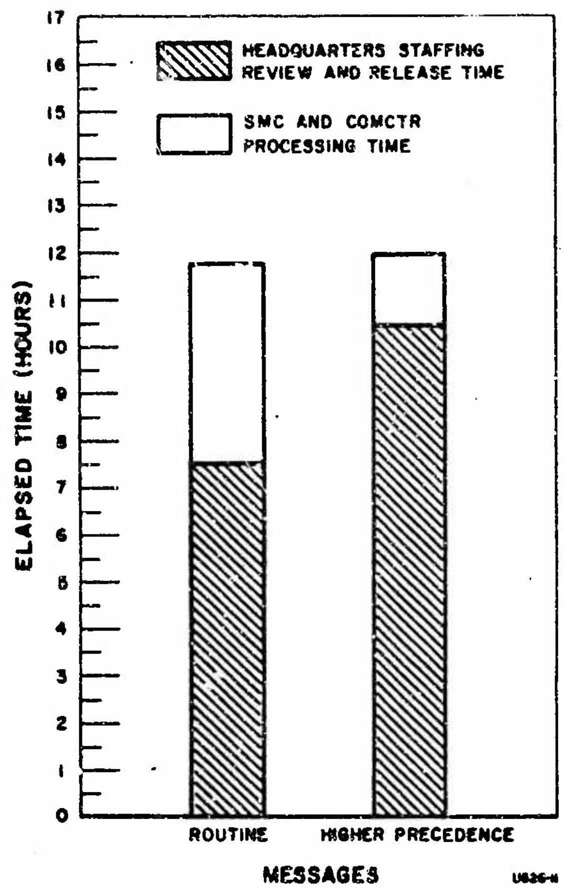


Figure 18. Average Headquarters Processing Time for Outgoing Routine and Higher Precedence Messages

THE U.S. ARMY SUPPLY AND MAINTENANCE COMMAND INVENTORY RESEARCH OFFICE

Mr. Bernard B. Rosenman
Inventory Research Office
SMC Systems Support Center
U.S. Army Supply and Maintenance Command

Presented at

Fifth U.S. Army Operations Research Symposium
Fort Monmouth, New Jersey
29 - 30 March 1966

14 February 1966

The U.S. Army Supply and Maintenance Command Inventory Research Office

Bernard B. Rosenman
Inventory Research Office
SMC Systems Support Center
Frankford Arsenal

The Inventory Research Office was established on 1 July 1965 under SMC General Order No. 59. Our charter calls for us to provide an in-house operations research capability, with primary emphasis on introducing the concepts of modern inventory theory into Army practice. The Office was formed from a small cadre of analysts that had begun work at Frankford Arsenal back in 1956 under the auspices of the U.S. Army Ordnance Corps Inventory Research Project and which has carried on a limited program of research and implementation from that time.

It has long been our feeling that operations research studies, particularly in the Army, have suffered in the translation from theory to practice because of the lack of an in-house capability that is able to perform the numerous and exacting tasks involved in the implementation process. The nature of the translation function is well understood in the case of hardware development and adequate provisions seem to have been made to carry on such activities as engineering test, product engineering, user test, field training and the like. Activities that are analogous to these must be carried on as a routine and continuing function in the operations research field as well, for it is quite unrealistic to expect that an operations research model, no matter how elegant, can be planted in the midst of an operating environment that is at best passive and find itself assimilated as a harmoniously functioning part of the host body. The task of finding or developing a model that has the possibility of functioning well, of adapting it to fit the host environment so that its procedures are compatible with others (for none of these models is global in scope), and of following through on the implementation to assure that the graft has actually taken is a long and arduous one. And, partly because intimate familiarity with the host environment is so important and because the time span from model formulation to completion of implementation is often long, it has been our feeling that this particular kind of operations research endeavor is best done by an in-house group. This, at any rate, has been the point of view we have espoused over the last several years and we have finally got the Supply and Maintenance Command to give it a try.

The work program of the IRO consists, roughly speaking, of three kinds of projects:

(1) Short-range "crash" jobs where immediate answers are needed.

(2) Longer-range projects involving some amount of model development or adaptation and significant amounts of pre-implementation and post-implementation activity.

(3) "Background" research where current state of the art is of little immediate value in so far as implementation programs are concerned.

All projects must be agreed to and funded by the Supply and Maintenance Command. Some originate with us; others (it should not be difficult to guess which kind) come from Command Headquarters. And all except the "background" research projects are rather well-defined, each with at least a few intermediate milestones that are understood by the headquarters people so that an acceptable degree of progress reporting can be done.

What we do specifically can best be explained by describing some of the projects we are currently carrying. It's perhaps best to start with the middle category--those projects that require some amount of model development and considerable amounts of pre- and post-implementation work--since projects of this kind are probably most familiar to this audience. The first I'd like to talk about is:

"A Management System for High Value Army Aviation Components"

A description of this system, the development of which was completed in July 1964, was given in a paper presented at last year's symposium.* The Supply and Maintenance Command and the Army Aviation Materiel Command agreed, in the spring of 1965 to conduct a test of the system on two Army aircraft systems beginning in July 1965. Our Office was given the task of developing and providing technical guidance of the test program.

It is of interest to note that, while there is a "model" in the usual operations research sense imbedded in this system, the most critical elements of the test program concern not the model but the information system within which the model must function. One of the specifications of the original study was that the new management system could impose no additional source data requirements on the Army information system. More specifically, we were asked to develop a system that

* Logistic Management of High-Value Aviation Components, by D. Hoekstra, presented at U.S. Army Operations Research Symposium, 30-31 March, 1 April 1965, Redstone Arsenal, Alabama.

could operate with data generated from the Army Equipment Record System (TAERS). This was done. But, during the course of development of the management system, many deficiencies were discovered in the TAERS source data. The system, therefore, provides for Regional Aircraft Logistics Managers in the various geographic areas where field maintenance and supply activities are carried on and for a man-machine transaction edit function at the Army Aviation Materiel Command (AVCOM). The effectiveness of their efforts in assuring a flow of TAERS data that can be used in the management of these high cost components is one of the most critical features to be measured in the test program.

One other major task in the test involves the processing of TAERS and related data at AVCOM to provide input data for the inventory model and for management of component assets. A set of computer programs was designed for collecting actuarial life histories, for measuring removal, repair, overhaul and other pipeline times and flow rates, and for showing present and projected item status. But AVCOM doesn't have programmers available to program this system nor has the expected augmentation of their computer equipment materialized for processing data generated from TAERS. Thus, the most essential phase of the test program---operating the proposed system with real data in the real operating environment---has not yet begun. We have, therefore, found it necessary to change the test plan. We have decided, in effect, to run a manual simulation of the computer program on the test items. This appears to be feasible since transaction volumes on the test aircraft are not large. There are, indeed, advantages to this approach that we had not originally thought of. The set of computer programs needed to implement the system is large and rather complex. Moreover, quality of the input data is poor and it is difficult to anticipate all of the problems that are likely to arise in routine processing. The manual simulation, it is hoped, will give us and AVCOM some badly needed experience in working with the real data and some insight into the types of problems the computer program will ultimately have to cope with. Thus, what started out as a desperation attempt to keep the test program moving appears to offer some very appealing features.

A second project being worked on involving a comparable amount of effort by our office but which is in an earlier stage of development is:

"Reparable Items Supply Control Models"

Fortuitously, a paper is being given at this symposium on this project* and further details do not have to be supplied by me except to point out that this is a good example of a project whose difficulty was badly under-estimated and which turned out to involve a great deal more theoretical development than expected.

*Supply Management Models for Reparable Items, by D. Hookstra, presented at Army Operations Research Symposium 29-30 March 1966, Ft. Monmouth, N. J.

This project grew out of contract work previously begun at MIT and University of Michigan. Three more of our intermediate-range projects have the same roots. One, recently completed, involved the adaptation of the MIT non-reparable model for use by the U.S. Army Logistical Center in Japan. This was rather a straight-forward job but does have the distinction of being one of the few "mail-order" operations research jobs we know of.

The USALC Japan asked us some time ago if we could help them apply statistical safety levels and economic ordering principles to their operation. We, by correspondence, outlined a data collection program for them for estimation of demand variability. They sent the data to us for analysis and, concurrently, conducted their own study on their costs of holding and requisitioning stock. They also sent us a description of their current operating rules and data on the composition of their catalog which were used to construct estimates of how the new system would compare in performance with their present system. Finally, they sent one of their men to our office last November, at which time we indoctrinated him in the features and methods of the new system. He then wrote the implementing procedures for the Center while at our office and, presumably, the system is now, or will shortly be, in operation.

Another involves the development of improved forecasting techniques for the Army's Inventory Control Points; the third in this category has to do with development of certain modifications to the MIT non-reparables model itself that operating experience indicates are needed. The forecasting project involves continuation of work begun at University of Michigan which we took over when their contract with the Army expired. A five-year history of individual demand and return transactions on a sample of about 60 Army Tank Automotive Center items has been collected and has been tested in an error-analysis simulator using various forms of moving average and exponential smoothing procedures. Analyses are also being made of Variance-to-mean Ratio estimation methods and of transaction order size distributions. All of their work involves analyses not only of the gross demand and return history, but also subsidiary analyses by type of demand (recurring, non-recurring) and type of customer (CONUS Army, MAP, Overseas, rebuild, etc.) The aim of the study is to develop improved forecasting methods suitable for routine computer application with a controlled "filter" for screening "outlier" transactions and a forecast error measurement procedure that will permit periodic adjustment of forecast model paramiters.

Work on modification of the MIT model has been started because of intuitively unsatisfactory results obtained in applying the model to certain classes of items. These difficulties are inherent in all "lot-size" optimization models and arise from the assumptions of cost

linearity and batch delivery. While we do have some ideas on how the model can be re-formulated, work has not yet progressed far enough to report on as yet.

The "background" research we are concerned with really involves only one area: multi-echelon inventory systems. We gave a paper on some of our work on this type of problem at last year's symposium.* Knowledge of the behavior of such systems has not advanced very much at all in the last year, either in theoretical work or in pragmatic applications. And, at least at the moment, the effort we are able to devote to further investigations in this area is minimal because of the press of other projects that are considered to be of higher priority. But we suspect that we may be able to devote resources to this research in the future because of growing Army realization that development of policies for individual organizational and functional segments of the logistic system is not the most satisfactory way to proceed.

One reason for our inability to devote more time to the multi-echelon problem (and, indeed, to some of the intermediate-range projects as well) has been the continuing influx of "crash" jobs. I am sure that all operations research organizations are familiar with these. Most of them are projects that would normally be done as staff studies or perhaps by inspection teams--and, as a matter of fact, it's not at all unusual to find the same project being worked on concurrently by non-operations research groups. It's perhaps stretching the point to call these operations research jobs since the time allowed for completion usually limits severely the amount of quantitative analysis that can be done. But there is no question that the background acquired in our other work is extremely helpful in "crash" jobs of this kind and we find them to be very fascinating and satisfying to work on. Some examples of these "crash" projects are:

A Method for Forecasting Army Stock Fund Sales

This had to be done within a very short period of time for use in the FY 1965 mid year budgetary reviews. It had been recognized that Army Stock Fund Sales are seasonal, tending to increase as the fiscal year progresses. No systematic technique for estimating seasonal coefficients was in use in the Army and, as a result, the mid-year reviews with OSD and Bureau of the Budget were difficult affairs. Our study involved really nothing more than application of standard time series techniques but it is heartening to be able to report that our recommendations were adopted and implemented in an AMC Regulation.

*Effect of Army Stockage Policies on Costs and Performance of the Supply System, by J. B. Henard, Jr. and B. Rosenman, presented at U.S. Army Operations Research Symposium, 30-31 March, 1 April 1965, Redstone Arsenal, Alabama.

A Quick Method for Estimating Funds Required to Support Southeast Asia Buildups

What was wanted here was a way to estimate repair part build-up requirements when specific information on the make-up of the forces to be re-deployed was lacking and when time did not permit querying the National Inventory Control Points for their estimates. The method we developed involved selection from the Troop Basis of combat type organizations that might be used as "building-blocks" of any force to be re-deployed, selection from their TO/Es of those major items of equipment that could be considered to be the heaviest parts users, and development of tables to be used in conjunction with a simple model for computing increased funds requirements resulting from pipe line extensions and increasing part failure rates. Then, if the question arose: Suppose we were to move "X" Infantry Brigades, "Y" Aviation Companies, "Z" Armored Divisions and their associated (but unspecified) support to VietNam by such and such a date, how much additional stock fund and PEMA money would be needed for parts support? people at Supply and Maintenance Command could compute by hand a somewhat defensible estimate within an hour. I don't think our method has been used as yet, for less formal estimating methods apparently still suffice, but it is available for use should the need arise.

Statistical Study of an NICP's Supply Control Operation

This project, which turned out to be a big resource consumer, resulted from the report of an inspection team's visits to one of the National Inventory Control Points. We were asked to determine if effects of migration of items between dollar value categories changes in assignments of analysts to items, policies on supervisory reviews and similar matters constituted serious management problems. Before the study got under way, we were also asked to examine in depth the supply management history of six critical items to determine if a discernable pattern of causes existed for their continuing backorder position. The first part of this study was primarily statistical and the data turned out to be of substantial interest to the NICP. As expected, the analyses of the critical items revealed no single pattern of causes for their continuing poor availability but we were able to devise a technique for isolating the major causes and for determining in a quantitative way the contribution of each cause to the backorders on hand at any given point in time. Although we hesitated to do so, we could not avoid making some recommendations on steps to be taken to improve the quality of supply control studies and, as a result, find ourselves involved in attempts to put some of them into effect.

The Inventory Research Office is, of course, too new to have a wealth of experience to report on and it is too early to tell whether or not an in-house organization of this kind can do an effective job in bridging the gap between research and implementation. Problems do exist, mainly with resources, that cause the work to move more slowly than we'd like. But we're hopeful that they can be overcome and that we'll have a good deal more finished work to report on at next year's meeting.

SUPPLY MANAGEMENT MODELS FOR REPAIRABLE ITEMS

Mr. Djoerd Hoekstra
Inventory Research Office
SMC Systems Support Center
U.S. Army Supply and Maintenance Command

Presented at
Fifth U.S. Army Operations Research Symposium
Fort Monmouth, New Jersey
29 - 30 March 1966

INTRODUCTION

This paper discusses various policies which can be followed by a National Inventory Control Point (NICP) in scheduling procurement and rebuild of repairable items. By and large, applications of inventory theory are still limited to non-repairable items; these items are bought by the Army NICP's in wholesale lots, sold to the Army retail supply system and not normally returned for repair when they fail, but discarded. It can be shown that for those non-repairable items, the NICP should establish a reorder point: when its on-hand and on-order assets fall below this reorder point, it should buy a given order quantity. Formulas have been developed to determine the "optimal" values to be used for the reorder point and for the order quantity, i.e., those values which will yield the best possible supply performance for a given investment in inventory on-hand and for a given procurement workload. These formulas are now widely used throughout the U.S. Army NICP's and equivalent wholesale organizations in the other services and the DSA. In supply management jargon, one uses the term Economic Order Quantity (EOQ) and Variable Safety Level (VSL) when referring to procurement lot sizes and reorder points thus computed.

Before the implementation of EOQ/VSL policies it was common practice to buy in lots which represented a fixed number of months (e.g. six months) of supply, without regard for the possibilities of trading off the cost of holding stocks against the cost of procurement actions. In the case of an inexpensive item for which there is only an occasional demand, it pays to buy several years worth of supply because the small added cost of stocking a greater quantity is more than offset by a reduction in the number of procurement actions. Conversely, in the case of an expensive item for which there is heavy demand, the economic scale tips in favor of lots which are as small as possible. The EQQ calculation considers the cost of holding a unit of stock and the cost of a procurement action, for each individual item, in arriving at the optimal buy quantity. The overall effect is a lower level of inventory (higher turnover rate) or a lighter workload of the procurement department or a combination of these two beneficial effects.

Before the introduction of Variable Safety Level calculations, the reorder point was normally set equal to the expected demand during the procurement lead time, plus a safety level again equal to a fixed number of months of supply. Thus, with a procurement lead time of six months, and a safety level of one month, the reorder point would equal seven months of supply for all items. The purpose of the safety level is to protect against running out of stock when the actual demand in the lead time exceeds the expected level. But, a fixed safety level of one month cannot be equally appropriate for all items. Ideally, the safety level should be set with regard to the demand variability which can be expected for a given item, and the desired level of protection against a stock-out. Under the Variable Safety Level scheme, each item will receive a different safety level which is designed to achieve the best possible level of customer service within the limits imposed by procurement workload and investment in inventory.

How to manage repairable items is much more difficult and still largely unsolved. These are the items which are not all discarded by the user in the field upon malfunction or failure, but of which some are returned to the NICP for overhaul or repair. They are typically the more complex and expensive components such as carburetors, starters, missile chassis, rangefinders, ballistic drives, and radios.

Replenishment of wholesale supplies can come from two sources: from new procurement and from repair of unserviceable items. This gives the decision maker an added degree of freedom in deciding when and how much to repair, considering the economic consequences of his decisions, and realizing that his repair decisions and procurement decisions should be coordinated to achieve the best possible supply performance. Repair as a source of supply differs from procurement in a number of ways. Repair is usually done Army depots, thus there is no contract to be negotiated with an outside manufacturer and consequently a much smaller set-up cost for a repair action than for a procurement action. Also, the lead time of a repair action is normally shorter than that for a procurement action. But the most fundamental difference between repair and procurement is that one can only repair as many items as there are unserviceable items available to be repaired.

There may, in some cases, be a question of the service expected from a repaired item versus the service expected from a newly procured item, but this question is outside the scope of this paper, where it is assumed that a repaired item is as good as a new one.

Why don't we yet have a uniform and theoretically justified policy for the management of repairable items similar to the EOQ/VSL policy? Certainly not because such a policy is considered unimportant: the dollar impact of improvements in the repair and procurement policies of these more expensive items are likely to be quite large compared with what has already been achieved in the case of the non-repairable items. The difficulty is two-fold. In the first place, there is no agreement on the form of the policy that should be followed. And secondly, even if one chooses one of the several policies which have been proposed, it is very difficult to calculate the optimal values of the various decision parameters involved.

Our research centers on two policies which differ widely in the degree of coordination which can be achieved between procurement and repair actions. In a way these two policies are extremes, and several others can be easily constructed by taking some features from the first, and others from the second. The U.S. Navy, for instance, is currently implementing a policy which is a cross-breed of the two policies discussed in this paper.

TWO STOCK POLICY

The first policy to be described was originated by Dr. Herbert P. Galliher, now at the University of Michigan, who was at MIT when this development started. It is named the Two Stock Policy, because it corresponds, conceptually, to a situation where there are two separate stock managers.

The simplifying assumption is made that we can split the total demand for serviceable items into two classes of customers, those who bring an unserviceable in exchange when they ask for a replacement item, and those who request a serviceable item without giving an unserviceable in exchange. The first class of customer goes to the first stock manager, whom we call the Repair Manager. He has two piles; for every unit he adds to the pile of unserviceable items, he loses one from the pile of serviceables. His total inventory, counting serviceables and unserviceables, remains always the same and all he does is to watch his pile of unserviceables grow and when it reaches a certain size, he repairs them. Supposedly he initiates the repair action early enough so that his chances of running out of serviceables while he waits for this order to be delivered from the repair shops are small.

The second manager only has to worry about procurement, and his actions are exactly the same as those of someone managing a non-repairable item. He watches his stock of serviceable items, and when it falls to the reorder point, he buys an economic lot from the manufacturer.

In reality, of course, we do not keep two separate piles of serviceable items, one of items to be given to those who return unserviceables, and one for empty-handed customers, and the independent actions of the hypothetical Repair and Procurement Managers can be restated as follows:

a. Establish a <u>reorder level</u> for procurement, an <u>economic procurement quantity</u> and an <u>economic repair quantity.</u>

b. Buy one procurement quantity when the level of serviceable items on-hand and on-order from procurement and repair, plus unserviceables on-hand falls below the reorder level.

c. Repair when the level of unserviceable items on-hand reaches the repair quantity.

This policy is simple and it has been possible to develop a mathematical model (Appendix A) which can be used to compute the "optimal" values of the three parameters (reorder level, procurement quantity, and repair quantity) for any item. However, the admirable simplicity of the Two Stock policy is at the same time its weak point. It does not achieve much coordination between repair and procurement actions. Repair is initiated

when a pre-established quantity of unserviceables has been accumulated, whether they are then needed or not. In some cases, when the supply of serviceables has dwindled unexpectedly, repair might be the proper action, even though less than the required number of unserviceables are on hand. In other cases, ample supplies to cover the needs of the foreseeable future may be on hand and repair might profitably be postponed, even though the required number of unserviceables are available.

In short, whether or not repair is initiated should depend on two considerations, the availability of unserviceables and the need for serviceables. The Two Stock Policy only considers the first factor, and it should therefore be possible to improve upon it.

A similar weakness exists with respect to procurement actions. In deciding whether or not procurement should be initiated, the Two Stock Policy will count all unserviceables on hand as assets against the reorder point without considering the extent to which repair actions can be counted upon to convert these unserviceables into serviceables.

LOOK AHEAD POLICY

The second policy overcomes these weaknesses but is thereby more complex. This policy has been appropriately named the Look Ahead Policy because it operates on estimates of future repair orders. To assure that repair actions are initiated when repaired items are needed to help prevent an out-of-stock position, a second reorder point is introduced. Just as the procurement reorder point warns the manager of the necessity of a "buy" action, so should the repair reorder point indicate when the level of serviceables on hand plus on order, due to arrive before the end of the repair lead time, is falling below the demand that can occur in the repair lead time. The repair reorder point triggers a repair action, but only if enough unserviceables are on hand to at least fill a minimum repair quantity.

Upon reaching the procurement reorder point, a decision has to be made whether procurement is required or not; a projection is made of the unserviceable items expected to be on hand when the repair reorder point is reached (it is assumed throughout this paper that the repair lead time is shorter than the procurement lead time).

Now the question is asked: is this projected quantity sufficiently large? The projection is to some extent a guess because it is always hard to predict when the repair reorder point will be triggered and how many unserviceables will be turned in during that period. By requiring that the projected repair order exceeds the minimum repair quantity by some protective level, we can limit the chances of putting false hopes on a repair action which later fails to materialize. If the projected repair order satisfies this requirement, it is then added to the level of serviceables on order as a planned repair order. If this brings the level of

serviceables on hand and on order above the procurement reorder point then no further action (procurement) is required. However, if either the projected repair quantity is too small, or the planned repair order is not large enough to bring serviceable assets above the procurement reorder point, then procurement action is taken.

Thus, the serviceable assets level, for the purpose of procurement decisions, consist of serviceables on hand (or minus the backorder position if the item is out of stock) plus serviceables due in from procurement and from repair plus any planned repair orders that may be on the books. Other than outstanding procurement orders, the planned repair orders are quite unreliable assets. They are counted upon to provide the necessary serviceables when needed as a substitute for procurement, but their real worth is uncertain: one large demand and they may turn out to be empty hopes; conversely, a rash of unserviceable returns may swell them to very substantial quantities. It is desirable, therefore, that they be continuously monitored and adjusted in size to always reflect the latest estimate of how many items will be available for repair and when. In addition, it is possible to "pad" the procurement reorder point if the asset level includes planned repair orders in compensation for their unreliability as assets. Another way of achieving the same end is to count the planned repair orders for something less than their expected size when calculating serviceable assets.

The Look Ahead Policy can be summarized as follows:

a. Establish two reorder points, one for procurement, the other for repair.

b. Establish an economic procurement quantity , a minimum repair quantity, and a minimum planned repair order.

c. When the level of serviceable assets falls below the procurement reorder point, determine whether the unserviceables expected to be on hand when the repair reorder point is reached is larger than the minimum planned repair order. Procure only if this is not so, or if the projected repair quantity is not large enough to bring assets above the reorder point.

d. Initiate a repair action for all unserviceables on hand when the level of serviceables on hand and on order - due in within the repair lead time - falls below the repair reorder point. Take no action, however, if the unserviceables number less than the minimum repair quantity.

Note that under this policy procurement decisions are not taken without consideration of the prospects for future repair actions, and repair actions are time phased to prevent stock-outs, within the limitations imposed by the availability of unserviceable items. Thus, the shortcomings of the Two Stock Policy have, in theory at least, been corrected. A mathematical model which enables one to calculate the "optimal" values for the five decision parameters in the Look Ahead Policy is under develop-

ment. The basic version of this model is given in Appendix B.

One of the interesting results obtained in Appendix B is the form of the solution for the optimal procurement quantity Q and procurement reorder level R.

$$Q = \sqrt{\frac{2\,\lambda_s\, j + 2(1+i)\int_R^{\infty}(y-R)\,G(y)\,dy}{i}}$$

$$\int_R^{\infty} G(y)\,dy = \frac{i}{1+i}\,Q$$

where λ_s = demand rate for serviceable items

$G(y)$ = probability that demand in procurement lead time exceeds y

i, j = cost factors

This solution is independent of the repairability of the item; in fact, this solution is identical to the solution for the non-repairable case. Procedures for calculating optimal procurement quantities and reorder points are now already in use at the NICP's and it would simply be a matter of extending these procedures to cover repairable items as well.*

* A few of the NICP's have already done this.

SIMULATION

We turned to simulation of these different policies, alongside a model of the policy presently being followed, initially for the purpose of performance evaluation .

Simulation programs were easily written in the special language for computer simulation, SIMSCRIPT, which has recently become available for most scientific computers. In the computer we can easily create the proper environment, with scanty data, improbable forecasts, changing requirements, uncertain deliveries, etc., under which these policies have to operate. Not that we have incorporated quite that much reality in our programs yet. Even in simulation models it is desirable to maintain a certain level of abstraction; "too much" reality can make the simulation model as difficult to study as the real problem and the model then loses a good deal of its usefulness. For instance, in our experiments thus far we have always used a stationary demand pattern, i.e., the mean demand rate and the average order size are kept constant during the run. But nevertheless a realistic amount of unpredictability is built into the demand and return processes by making demand and return transactions occur at random time intervals, and by making the number of items involved in individual demand and return transactions follow a geometric probability distribution which exhibits considerable variability. This way the demand can be made as erratic as required for correspondence with actual experience on similar items.

The question that we originally set out to answer by simulation was whether the improvement in performance of the Look-Ahead Policy over the Two-Stock Policy is realizable, or whether the uncertainties in demand and return rates, and perhaps in lead times, make the added considerations of no value. Suspecting that the Look Ahead Policy is superior, the intent was to find out whether this could actually be demonstrated under circumstances which approximate reality.

However, this original objective decreased in importance as soon as we began to run the simulations, for confronted with the actual achievements of the simulated policy, we were forced to reexamine assumptions, find solutions for unforeseen circumstances, and, in general, reduce the original theory to a set of consistent and practical rules which achieves the desired performance. In the case of the Look Ahead Policy, the very first runs did indeed show it to be slightly better than the Two-Stock Policy in actual performance, but at the same time there was very little correspondence between what was happening in the simulation and what was supposed to have happened according to our mathematics. For instance, we originally assumed that the average repair quantity depended on the procurement quantity and the minimum planned repair order as follows:

$$\bar{R} = Q_R + 1/2 \frac{\lambda_R}{\lambda_S} Q_P$$

where $\bar{R}$ = average repair quantity

Q_R = minimum planned repair order

Q_P = procurement quantity

In our conception, the level of unserviceables on hand and due in would build up with successive procurements, each procurement postponing repair for another period averaging Q_P/λ_S, which means on the average $\lambda_R Q_P/\lambda_S$ unserviceables gained, until at least a quantity Q_R had been reached. Q_R would thus be "overshot" by $1/2\ \lambda_R Q_P/\lambda_S$ on the average. However, it soon turned out that this is true only if the amount of unserviceables accumulated during a procurement cycle $\lambda_R Q_P/\lambda_S$ is small compared to the number of unserviceable items needed to make up a planned repair order, Q_R. With a large fraction of items returned for repair, $\lambda_R/\lambda_S = .80$ for instance, $\lambda_R Q_P/\lambda_S$ tends to be as large as or larger than Q_R which exactly reverses the situation. Instead of having a sequence of procurement orders followed by one repair order, we then observe only one procurement followed by a sequence of repair orders. The single procurement cycle is then usually long enough to accumulate enough unserviceables (more than Q_R) for a planned repair order. This led us to the development of a second formula for the average size of repair orders:

$$\bar{R} = \frac{2\lambda_R}{\lambda_R+\lambda_S} Q_R + \frac{\lambda_R}{\lambda_S} Q_P \quad , \text{ if } \frac{\lambda_R}{\lambda_S} Q_P > Q_R$$

The first term represents the repair order projection (less than Q_R and therefore rejected) made at the time a procurement order is placed, and the second term represents the additional unserviceables which accumulate during the procurement cycle (until the procurement reorder level is again triggered).

As far as the dependence of $\bar{R}$ on Q_R is concerned, this hardly differs from the earlier expression which contained the term Q_R because $2\lambda_R/\lambda_R + \lambda_S$ would be close to one. But the effect of Q_P on $\bar{R}$ is now twice as large with λ_R/λ_S instead of $1/2\ \lambda_R/\lambda_S$. In Appendix B the simplifying assumption was made that the expression can be written as

$$\bar{R} = \alpha Q_P + \beta Q_R$$

where α and β are constants not depending on Q_P or Q_R. The optimal procurement quantity Q_P and the optimal average repair quantity R are shown not to depend on α or β. However, if one wants to find the correct Q_R to be used with the optimal Q_P to produce the desired $\bar{R}$, an estimate of α and β has to be available.

Other things learned in observing the simulated model were:

a. It pays to discount somewhat the assets under the Look Ahead Policy if planned repair orders are involved because we then count on future returns of unserviceable items; this adds to the uncertainty of the situation which otherwise would consist of demand variability only.

b. Demands and returns being as erratic as they are, actual repair orders under the Look Ahead Policy are not of approximately the same size each time as is the case in the Two Stock Policy but fluctuate considerably. This results in an average level of stock on hand larger than indicated by the expression used in Appendix B.

c. Because demand transactions in many cases request more than one item at the time, one expects to "undershoot" the reorder level used. After having calculated the optimal reorder level to be adhered to one should add to this the expected undershoot or else the effective level of protection against stock-out may be much less than expected. This much was known and one normally estimates the expected undershoot as one-half times the average demand order size. The simulation runs showed, however, that the procurement reorder level undershoot under the Look Ahead Policy is much larger than one-half the average demand order size, amounting in many cases to several times the average order. The reason was found to lie in the planned repair orders; when a demand occurs they are adjusted downwards causing the asset level to drop by more than just the quantity demanded.

CONCLUSIONS

The major conclusion we have drawn from our work to date is that if one deviates from a simple policy with a nice clean mathematical solution, such as the Two Stock Policy, be prepared for trouble. If, knowing this, one experiments anyway because there appears to be the possibility of significant improvement, then simulation is an absolute necessity. It is hard to conceive of a decision rule which is so simple that nothing can be gained from simulation experiments. And with special computer languages such as SIMSCRIPT, simulation has suddenly become easy to use and economical, where previously it was difficult, expensive and, above all, time consuming.

All the problems associated with the Look Ahead Policy have not yet been solved to our satisfaction. However, even in its current, admittedly imperfect, form, the Look Ahead Policy has in some simulation runs shown itself to be better than the Two Stock Policy. But then, those runs may have painted an unfair picture. It should be pointed out that in simulation runs so far the (programmed) decision maker is assumed to know the true value of demand and return rates, probability distributions of order sizes, and lead times. This is a welcome simplification; it means, for instance, that the various decision parameters such as lot sizes and reorder levels have to be calculated only once at the beginning of the run. However, it eliminates a good deal of actual uncertainty, and it may well give the Look Ahead Policy an unfair advantage. We plan, therefore, to incorporate a forecast of future demands and returns based on past experience, with periodic recalculations of the decision levels before drawing any conclusions as to the superiority of one policy over the other.

REFERENCES

1. FRANKFORD ARSENAL: Proceedings of Seminar on Advanced Inventory Concepts, held at Massachusetts Institute of Technology, June 1962

2. GALLIHER, H.P.: Replenishment of National Stock by Procurement and Rebuild, (unpublished notes)

3. GALLIHER, H.P., Philip M. Morse, and M. Simond: Dynamics of Two Classes of Continuous Review Inventory Systems, Journal of the Operations Research Society of America, Vol. 7, no. 3 (June 1959)

4. PHELPS, E.S.: Optimal Decision Rules for the Procurement, Repair, of Disposal of Spare Parts, RAND Corporation Memorandum RM-2920-PR, May 1962

5. PRITCHARD, J.W.: Inventory Model for Repairable Items - Theory and Practice; paper presented at the March 26, 1964 meeting of the Philadelphia Chapter of the Operations Research Society of America

APPENDIX A

The Two-Stock Policy

In conformance with the earlier part of this paper, a distinction is made, at least conceptually, between a Repair Manager who serves those customers returning unserviceable items to cover their demands for serviceable items, and a Procurement Manager who serves those customers not giving unserviceable items in trade. The subscript 1 in this Appendix refers to the Procurement Manager's business, the subscript 2 indicates the Repair operation.

State Probabilities

First consider the Repair Manager. His total stock, serviceables and unserviceables, always remains at the same level, because for each serviceable item given out, an unserviceable item is received. We denote this total level by M_2. By repairing each time the unserviceables on hand reach Q_2, the repair quantity, the level of serviceable assets A_2 (on hand minus backorders plus on order from repair) will always stay between M_2 and M_2-Q_2. It can be shown that at any given time A_2 can be anywhere in the interval with equal probability; A_2 is rectangularly distributed in the interval between M_2 and M_2-Q_2 with density function $f(A_2) = 1/Q_2$.

There is a direct relationship between the serviceable assets at time t and the net stock level (on hand minus backorders) a full repair lead time later. Since the on order portion of the asset level will be delivered before the end of the lead time, the net stock position at the end of the lead time must be equal to the earlier asset level minus the demand that occurred during the lead time.

$$N_2\ (t+T_2) = A_2(t) - Y_2$$

where T_2 is the repair lead time, and Y_2 is a random variable representing the demand on the Repair Manager during a repair lead time.

This same equation can be restated, somewhat more conveniently, relative to M_2 as:

$$S_2\ (t+T_2) = H_2(t) + Y_2$$

where H_2 represents the amount by which the asset level is below M_2, and S_2 is the difference between M_2 and N_2.

We want to find the probability distribution of S_2. Any given S_2 is the sum of two random variables, a level H_2 some time T_2 in the past, rectangularly distributed between 0 and Q_2, and a demand variable Y_2 which we assume has a density function $g_2(y)$. Consequently, we find the state probabilities of S_2, $p_2(x)$ as:

$$p_2(x) = \frac{1}{Q_2}\int_0^{Q_2} g_2(x-t)\,dt$$

$$= \frac{1}{Q_2}\left[F_2(x) - F_2(x-Q_2)\right] \quad \text{-----} (1)$$

where $F_2(x)$ = the probability that Y_2 is less than or equal to x.

In the case of the Procurement Manager, the serviceable asset level A_1 (on hand minus backorders plus on order from procurement) will stay between the stockage objective M_1 and the reorder level $M_1 - Q_1$, and in probabilistic terms is rectangularly distributed with density funtions $1/Q_1$. Slack stock S_1 (the difference between M_1 and net stock level N_1) again is composed of a rectangularly distributed variable $H_1 = M_1 - A_1$ and a stochastic variable Y_1 representing the demand upon the Procurement Manager during a full procurement lead time T_1. Completely analogous to (1) we now find

$$p_1(x) = \frac{1}{Q_1}\left[F_1(x) - F_1(x-Q_1)\right] \quad \text{---} (2)$$

However, our final objective is to find the state probabilities of total net stock for the Repair and Procurement Manager combined because this enables us to calculate the average level of backorders and stock on hand. Total net stock N equals $N_1 + N_2$. We can also define a total stockage objective M equal to $M_1 + M_2$ and instead of the state probabilities of N we could just as well use the state probabilities of a slack stock S defined as M - N. It can easily be shown that $S = S_1 + S_2$. S is thus the sum of two independent stochastic variables S_1 and S_2 with known state probabilities $p_1(x)$ and $p_2(x)$. This leads to

$$\text{prob.}(S=x) = p(x) = \frac{1}{Q_1 Q_2}\int_{x-Q_1}^{x}\left[F(y) - F(y-Q_2)\right]dy \quad \text{---} (3a)$$

$$= \frac{1}{Q_1 Q_2}\int_{x-Q_2}^{x}\left[F(y) - F(y-Q_1)\right]dy \quad \text{....} (3b)$$

where $F(y)$ is defined as the probability that the sum of the demand on the Repair Manager during the repair lead time, Y_2, and the demand on the Procurement Manager during the procurement lead time, Y_1, is less than or equal to y.

The state probabilities of slack stock S tell us all we need to know. The probability of S = M, for instance, is equal to the fraction of time that we have exactly zero net stock; S greater than M means backorders, and therefore the sum of state probabilities for S equal to M + 1, M + 2, M + 3, etc. is equivalent to the fraction of time the system is backordered. The average level of backorders equals:

$$\bar{B} = \int_M^\infty (M - x)\, p(x)\, dx \qquad \text{---- (4)}$$

The average net stock level is found as:

$$\bar{N} = \int_0^\infty (M - x)\, p(x)\, dx \qquad \text{---- (5)}$$

and average stock on hand:

$$\bar{I} = \int_0^M (M - x)\, p(x)\, dx$$

$$= \bar{N} + \bar{B} \qquad \text{---- (6)}$$

Cost Minimization

The objective function to be minimized takes the form of an expression of Total Variable Cost:

$$TVC = C_b \bar{B} + C_i \bar{I} + C_u \bar{U} + C_p f_p + C_r f_r \qquad \text{---- (7)}$$

where

$\bar{B}$	=	average backorder level
$\bar{I}$	=	average level of serviceable stock on hand
$\bar{U}$	=	average level of unserviceable stock on hand = 1/2 Q_2
f_p	=	frequency of procurement actions
f_r	=	frequency of repair actions
C_i	=	cost of holding one unit of serviceable stock for one year
C_u	=	cost of holding one unit of unserviceable stock for one year
C_b	=	penalty cost for stock-out per backorder year
C_p	=	set-up cost incurred with each procurement ~~action~~
C_r	=	set-up cost incurred with each repair action

This total cost expression is a function of M, Q_1, and Q_2. The conditions for a minimum cost solution are found by taking the partial derivatives of (7) with respect to the three decision parameters and setting these three partial derivatives equal to zero, resulting in:

$$-\frac{i}{Q_1}\bar{I} + i\,V_1 + \frac{1}{Q_1}\bar{B} - V_2 - \frac{k\,\mu_p}{Q_1^2} = 0 \qquad \text{---- (8)}$$

$$-\frac{i}{Q_2}\bar{I} + i\,V_3 + \frac{1}{Q_2}\bar{B} - V_4 - \frac{j\,\mu_r}{Q_2^2} + \frac{i}{2} = 0 \qquad \text{---- (9)}$$

$$\int_M^\infty p(x)\,dx = i/(1+i) \qquad \text{----(10)}$$

where:

$$i = C_i / C_b = C_u / C_b$$

$$j = C_r / C_b$$

$$k = C_p / C_b$$

and:

μ_p = expected demand per year on Procurement Manager

μ_r = expected demand per year on Repair Manager

Furthermore:

$$V_1 = \frac{1}{Q_1 Q_2} \int_0^M (M - x) \left[F(x - Q_1) - F(x - Q_1 - Q_2)\right] dx$$

$$V_2 = \frac{1}{Q_1 Q_2} \int_M^\infty (M - x) \left[F(x - Q_1) - F(x - Q_1 - Q_2)\right] dx$$

$$V_3 = \frac{1}{Q_1 Q_2} \int_0^M (M - x) \left[F(x - Q_2) - F(x - Q_1 - Q_2)\right] dx$$

$$V_4 = \frac{1}{Q_1 Q_2} \int_M^\infty (M - x) \left[F(x - Q_2) - F(x - Q_1 - Q_2)\right] dx$$

The problem now becomes one of numerical analysis. We have three equations (8), (9), and (10) which have to be solved to find the "optimal" values of the three unknowns Q_1, Q_2, and M. The most practical way to do this is by iterating between M on the one hand and Q_1, Q_2 on the other. Given a starting value of Q_1 and Q_2 we can calculate M from equation (10); we start with a very large value for M, and keep decreasing it in small steps until (10) is satisfied. Then, using this value for M we go to equations (8) and (9) to solve for Q_1 and Q_2; this requires a complicated scheme because $\bar{B}$, $\bar{I}$, V_1, V_2, V_3, and V_4 are not really constants but depend on Q_1 and Q_2. Convergence takes place most rapidly if a "gradient" method is employed:

1. For given starting values of Q_1 and Q_2 evaluate the left hand side of (8) which we call $\mathfrak{J}$, and the left hand side of (9), η. If both $\mathfrak{J}$ and η are zero then the equations are satisfied and the present values of Q_1 and Q_2 are optimal if used with the present value of M.

2. If $\mathfrak{J}$ or η are not zero, determine the rates of change of $\mathfrak{J}$ and η with changes in Q_1 and Q_2,

$$\partial\mathfrak{J}/\partial Q_1 \quad \partial\mathfrak{J}/\partial Q_2 \quad \partial\eta/\partial Q_1 \quad \partial\eta/\partial Q_2$$

3. To find the changes to be made in Q_1 and Q_2 which will bring $\mathfrak{J}$ and η as close as possible to zero, solve the following two equations for ΔQ_1 and ΔQ_2:

$$\Delta Q_1 = \left(-\frac{\partial\eta}{\partial Q_2}\mathfrak{J} + \frac{\partial\mathfrak{J}}{\partial Q_2}\eta\right) \Big/ \left(\frac{\partial\eta}{\partial Q_2}\cdot\frac{\partial\mathfrak{J}}{\partial Q_1} - \frac{\partial\mathfrak{J}}{\partial Q_2}\cdot\frac{\partial\eta}{\partial Q_1}\right)$$

$$\Delta Q_2 = \left(-\mathfrak{J} - \frac{\partial\mathfrak{J}}{\partial Q_1}\Delta Q_1\right) \Big/ \frac{\partial\mathfrak{J}}{\partial Q_2}$$

4. Find a new Q_1 = old Q_1 + ΔQ_1

   and a new Q_2 = old Q_2 + ΔQ_2
   Go back to Step 1.

Having found the value of Q_1 and Q_2 satisfying (8) and (9) we can now go back to (10) and recalculate M. Then go back and recalculate Q_1 and Q_2. Keep iterating between M and Q_1, Q_2 until the solution converges.

APPENDIX B

THE LOOK AHEAD POLICY

Unfortunately, a simple approach using state probabilities is not available in this case. Instead, we have formulated approximate expressions for average backorders, average net stock, frequency of procurement and repair actions. The decision parameters which we want to determine are:

Q_P : procurement quantity

R_P : procurement reorder level

Q_R : minimum planned repair order

R_R : repair reorder level

The objective is to minimize Total Variable Costs:

$$TVC = C_b\bar{B} + C_i\bar{I} + C_u\bar{U} + C_p f_p + C_r f_r \quad --- \quad (1)$$

The symbols used here are the same as in Appendix A. A few more terms have to be defined:

λ_R = average return rate for unserviceable items

λ_S = average demand rate for serviceable items

$\bar{R}$ = average repair quantity

In the discussion on page 8 of this paper we have already exp ed that the average repair quantity $\bar{R}$ depends on both Q_P and Q_R in a rather complex manner, but for purposes of optimization we will assume the expression to be linear:

$$\bar{R} = \alpha Q_P + \beta Q_R \quad --- \quad (2)$$

The frequency of procurement orders is found as:

$$f_P = (\lambda_S - \lambda_R) / Q_P \quad --- \quad (3)$$

and the frequency of repair orders as:

$$f_R = \lambda_R / \bar{R} \quad --- \quad (4)$$

Furthermore average serviceable stock on hand is equal to average net stock plus average backorders.

$$\bar{I} = \bar{N} + \bar{B} \quad --- \quad (5)$$

Unserviceables on hand drop to zero each time a repair order is initiated, and then build up again to a repair quantity in sawtooth fashion. If all repair orders are of equal size, the average number of unserviceables on hand is then equal to one-half the repair order quantity as is the case with the Two Stock policy. In this case repair orders are not always of the same size, but we assume that for purposes of optimization we can still use the approximation

$$\bar{U} = 1/2\ \bar{R} \qquad --- \quad (6)$$

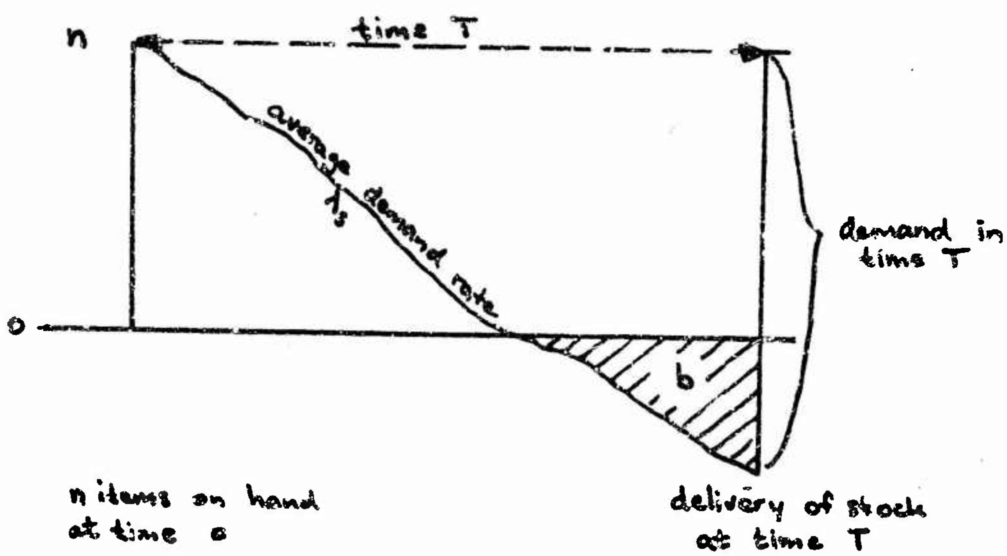
The only two levels for which an expression is not readily available are average backorders $\bar{B}$ and average net serviceable stock $\bar{N}$. The following two sections are devoted to the derivation of these expressions.

Calculation of Average Backorders B

Consider a situation with n serviceable items on hand and a time period T to go before the next delivery of serviceable stock will be made. It can be shown that the expected value of the backorder area (cross hatched) is equal to

$$b = \frac{1}{\lambda s} \int_n^\infty (y-n)\ G(y : T)\ dy$$

where $G(y:T)$ is the probability that the demand during the period T exceeds y



We can apply this formula to both the procurement reorder level and to the repair reorder level. First consider the case where a procurement order is placed upon reaching the procurement reorder level R_P. It will take a period T_P before this order is delivered and in the mean time we expect a backorder area (measured in backorder years) of

$$b_P = \frac{1}{\lambda S} \int_{R_P}^{\infty} (y - R_P) \; G(y;\; T_P) \; dy \qquad --- (7)$$

This situation arises f_P times per year on the average.

Similarly in the case where no procurement is initiated upon reaching R_P, but a repair order is subsequently placed where R_R is reached, we expect

$$b_R = \frac{1}{\lambda S} \int_{R_R}^{\infty} (y - R_R) \; G(y;\; T_R) \; dy \qquad --- (8)$$

This situation arises f_R times per year.

The total expected backorder area per year is thus found as:

$$\bar{B} = f_P b_P + f_R b_R \qquad --- (9)$$

Calculation of Average Net Serviceable Stock $\bar{N}$

Suppose that one would procure a quantity Q_P when the level of serviceables on hand and due in reaches the procurement reorder level R_P. After the procurement lead time T_P but just before the procurement order arrives we expect to have an amount of stock on hand equal to the Reorder level minus the expected lead time demand. Then the net stock level shoots up over a distance Q_P upon the arrival of the order, followed by a decline until the next delivery. Assuming that this next delivery is again a procurement, so that the expected value of net stock just before its arrival is again equal to $R_P - \lambda_S T_P$, then the average level of net stock in the interval between deliveries is:

$$n_P = R_P - \lambda_S T_P + 1/2 \; Q_P \qquad --- (10)$$

Similarly, if a system were subjected to repair deliveries only, and all deliveries are of the same size $\bar{R}$, then, the net stock level equals

$$n_R = R_R - \lambda_S T_R + 1/2 \; \bar{R} \qquad --- (11)$$

The system is under the influence of n_P for a fraction of time $(\lambda_S-\lambda_R)/\lambda_S$ and under the influence of n_R for a fraction λ_R/λ_S. Consequently, the overall net stock level is approximately

$$\bar{N} = \frac{\lambda_S-\lambda_R}{\lambda_S} n_P + \frac{\lambda_R}{\lambda_S} n_R \quad --- \quad (12)$$

Cost Minimization

Using the above approximations for $\bar{B},\bar{I},\bar{U},f_P$, and f_R, we now differentiate the Total Variable Cost expression (1) with respect to Q_P, R_P, Q_R, and R_R and set the four partial derivatives equal to zero. Assuming furthermore that the cost of holding unserviceable items, C_u, equals the cost of holding serviceable items, C_i, we obtain:

$$Q_P = \sqrt{\frac{2\lambda_S[j+(1+i)b_P]}{i}} \quad ---- \quad (13)$$

$$\bar{R} = \sqrt{\frac{2\lambda_R\lambda_S[k+(1+i)b_R]}{i(\lambda_R+\lambda_S)}} \quad ---- \quad (14)$$

$$\int_{R_P}^{\infty} G(y;T_P)\,dy = \frac{i}{1+i} Q_P \quad ---- \quad (15)$$

$$\int_{R_R}^{\infty} G(y;T_R)\,dy = \frac{i}{1+i} \bar{R} \quad ---- \quad (16)$$

where : $i = C_i/C_b$

$j = C_p/C_b$

$k = C_r/C_b$

Q_P and R_P can be numerically obtained by iterating between (13) and (15). Similarly (14) and (16) can quite easily be solved for $\bar{R}$ and R_R.